

Rowan University

Rowan Digital Works

Theses and Dissertations

4-8-2020

Teachers' sensemaking: Middle school and high school language arts grading and assessing practices for writing

Lana Michele Cook
Rowan University

Follow this and additional works at: <https://rdw.rowan.edu/etd>



Part of the [Language and Literacy Education Commons](#), and the [Secondary Education Commons](#)

Recommended Citation

Cook, Lana Michele, "Teachers' sensemaking: Middle school and high school language arts grading and assessing practices for writing" (2020). *Theses and Dissertations*. 2771.
<https://rdw.rowan.edu/etd/2771>

This Dissertation is brought to you for free and open access by Rowan Digital Works. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of Rowan Digital Works. For more information, please contact graduateresearch@rowan.edu.

**TEACHERS' SENSEMAKING: MIDDLE SCHOOL AND HIGH SCHOOL
LANGUAGE ARTS GRADING AND ASSESSING PRACTICES FOR WRITING**

by

Lana Michele Cook

A Dissertation

Submitted to the
Department of Educational Leadership
College of Education
In partial fulfillment of the requirement
For the degree of
Doctor of Education
at
Rowan University
March 25, 2020

Dissertation Chair: Monica Reid Kerrigan, Ph.D.

© 2020 Lana Michele Cook

Dedication

I dedicate this dissertation, and all of my life's work promoting the education and well-being of children, to my sons: James Richard, Robert William, and Daniel Henry. My three children are entrusted to the very system I aspire to impact, and as a classroom teacher, this knowledge drives my daily work. Nothing has made my work feel more urgent, important, or necessary than my role as a mother; the lens through which I view my classroom and the world is only possible because of the love I hold for each of you. Thank you for your patience with my work and for the myriad ways in which you inspire me every single day.

Acknowledgments

This work was patiently and thoroughly guided by my committee chair; thank you, Dr. Monica Kerrigan. Thank you also to Dr. Sam and Dr. Mitani for your feedback and advice. Phil, Bobbie, Amy, and Erin, and the Rowan University fall 2015 online cohort, thank you for your ongoing support, intelligence, humor, and advice. Thank you to all of the educators who so graciously dedicated their time to participate in these interviews and to the beloved schools I have taught in, particularly Haddonfield Friends School and my current teaching community. I would also like to acknowledge my parents, Rick and Bettyanne Murray. You created a support system of love and advice that has carried me through adulthood. Mom, you have inspired me as a mother, an educator, and a believer in the limitless promise and potential of all children. You taught me that adults have the power to shape children, and that we always need to wield that power with patience, love, and altruism. Dad, you taught me endurance and fortitude-- to demand more of myself and to forge ahead, regardless of what lay before me. To my siblings and their amazing families: Rick, Michele, Mikayla, Bryanna, Richard, Tara, Kevin, Aidan, Carys, Casey, Sean, and Annelien, I am so grateful for each of you. To my amazing sons, who reminded me time and again to be “perseverant.” I know that your future will be better for the insights I have gained in this process. Finally, to my husband, James, for the better half of my life, you have been by my side. Thank you for your unconditional support, encouragement, and love; thank you for pushing me to be and do my best. I love you all more than words could express.

Abstract

Lana M. Cook

TEACHERS' SENSEMAKING: MIDDLE SCHOOL AND HIGH SCHOOL GRADING
AND ASSESSING PRACTICES FOR WRITING
2019-2020

Monica Reid Kerrigan, Ph.D.

Doctor of Education

This qualitative case study will examine how middle and high school language arts teachers in a single school district make sense of their grading practices. This paper explores how Language Arts teachers at different grade levels may be faced with a variety of contextual factors that influence their grading practices. In order to do justice to this topic, a literature review will situate and contextualize writing instruction, learning standards, and assessments. Using a qualitative single case study design, this study will present the findings from nineteen in-depth teacher interviews, document analysis, and field notes.

Table of Contents

Abstract.....	v
List of Tables.....	xi
Chapter 1: Introduction.....	1
Problem Statement.....	2
Purpose of the Study.....	3
Literature Overview.....	4
Rationale and Significance.....	6
Theoretical Framework.....	6
Methodology.....	8
Summary.....	8
Definition of Key Terms.....	8
Chapter 2: Literature Review.....	10
Literature Search Strategy.....	11
Language Arts Writing and Instructional Models.....	13
Writing Instruction.....	14
Workshop and Traditional Models of Instruction.....	15
Grading.....	16
Assessment.....	17
A History of Grading.....	18
A Brief Taxonomy of Grading Schemas.....	19
Teachers' Values and Beliefs.....	20
Horizontal and Vertical Misalignment.....	25

Table of Contents (Continued)

Validity.....	29
Best Practice	32
Start with Clear Goals.....	33
Feedback.....	33
Accuracy	34
Misalignment With Best Practice	35
Theoretical Framework.....	38
Summary	41
Chapter 3: Methodology	42
Research Questions.....	42
Research Design	43
Setting	44
The School District	45
Sample and Participant Selection	45
Data Collection.....	46
Interviews	47
Documents.....	48
Field Notes	48
Research Validity	49
Credibility.....	50
Reliability	50
Data Analysis	51

Table of Contents (Continued)

Ethical Considerations	55
Institutional Review Board.....	57
Limitations	58
Summary	58
Chapter 4: Findings	60
Participant Sample and Setting.....	61
Data Analysis	62
Themes in the Data	63
Time	64
Perceiving Limits with Timeliness	66
Perceiving Limits: Doing Justice.....	67
Mitigating Time Constraints.....	68
Considering Motivation	71
Perceptions of Motivating Factors	71
Perceptions of Safeguarding Students.....	73
Setting Expectations Before Grading	74
Setting Expectations.....	74
Expectations and the Instructional Age Group.....	76
Offering Feedback	78
Estimating the Value of Feedback	78
Lean Feedback	79
Responding to Context.....	86

Table of Contents (Continued)

Perceiving Alignment.....	88
Perceiving Autonomy.....	89
Summary	90
Chapter 5: Discussion, Conclusions, Recommendations	92
Contextual Factors	93
Time	94
Discussion of Findings.....	99
Teachers Values and Beliefs	100
Inconsistent Use of Assessment Tools.....	100
Discussion of Findings.....	105
Making Sense: Grading Writing a Social Context	106
Time as an Investment: Measuring the Returns on Feedback.....	106
Discussion of Findings.....	107
Teachers Experiences in Their Classrooms	108
Assessing Reading Through Writing	108
Teaching to a Range of Students	109
Discussion of Findings.....	109
Limitations	110
Recommendations.....	110
Recommendations for Future Research	111
Recommendations for District Policy.....	112
Recommendations for Educational Leaders.....	113

Table of Contents (Continued)

Implications for my Teaching Practice	114
Conclusion.....	115
References.....	116
Appendix A: Qualitative Informed Consent	126
Appendix B: Interview Log and Introductory Comments	128
Appendix C: Alignment Matrix	129
Appendix D: Interview Protocol	131
Appendix E: Document Protocol	133

List of Tables

Table	Page
Table 1. Participant Demographics	62
Table 2. Frequently Used Codes During Initial Coding	63
Table 3. Pattern Coding, Themes	64

Chapter 1

Introduction

From an early age, American children are indoctrinated to evaluation in schooling, from assessment, to feedback, to grades. Yet, grading in America has become a controversial topic, eliciting a variety of stakeholder perspectives. Twitter, the prolific social networking platform—and host to active and ongoing social dialogue—has seen the spawn of this sometimes fractious debate, with the emerging conversations ranging from polemical (#ungrading, #AbolishGrading) to constructive dialogues promoting an adoption of research-based best-practices in evaluation. Leading experts on educational assessment advocate implementation of effective evaluation systems, insisting that the problem is not the evaluative aspects of grading, but rather the failure to implement effective systems and practices. A leading voice in assessment and grading, Dr. Susan Brookhart, describes the problem as the “...absence of implementation despite many assessment systems which have been written and developed over the years” (Brookhart, Stiggins, McTighe, & William, 2019). Another leading expert on grading, Dr. Dylan Williams, insists that grading is crucial, as measures of learning are essential to education (Brookhart et al., 2019). Given the expert consensus on the value of grades—when done right—the conversation to which this research seeks to contribute is not one on whether or not grades should be abolished but one on how we can understand the systems and practices as they exist. This research will alternately explore teachers’ sensemaking of their grading and assessing practices and the potential disjuncture between teachers’ practices in individual classrooms and between grade levels.

Each chapter in this dissertation serves a specific purpose. This chapter will set the stage for the discussion to follow. This chapter includes an overview of the topic, the problem and purpose for the research, an overview of the literature and theoretical framework, the significance of the study, a brief overview of the methodology, and a definition of key terms. This chapter concludes with a summary. Chapter two highlights many themes that emerged from the literature, including “hodgepodge” grades (Brookhart, 1991) and the influence of teachers’ values and beliefs on their grading practices. Chapter two will also outline sensemaking (Weick, Sutcliffe, & Obstfeld, 2005) as the theoretical frame. Chapter three will explain why a qualitative case study was used, and how the data presented was coded, analyzed, and triangulated. Chapter four will present the results of the research study, by offering the data collected, and chapter five will present the findings, conclusions, and recommendations of this research.

Problem Statement

Teachers often adopt idiosyncratic practices when it comes to grading (McMillan, 2001) and these idiosyncratic practices leave potential for differences in teachers’ grading practices, horizontally (amongst subjects in a single grade), as well as vertically (from one grade to another). Grades can affect students’ motivation, coursework track, and future prospects, and differences in teachers’ grading practices may be disruptive or problematic therein. Differences in teachers’ grading and assessing practices may alter the messages being communicated with regard to a students’ academic performance. This study is primarily interested in exploring teachers’ sensemaking with regard to the grading and assessment of students’ writing. Though writing is only one element of grading in language arts, it is an important one. Many college-bound high school

graduates face the prospect of developmental coursework in reading, writing, or math (U.S. Department of Education, 2017). This study is concerned with exploring teachers' sensemaking with regard to writing instruction, with the understanding that students' educational trajectory and future prospects are likely impacted by the grades that they receive in this area.

Purpose of This Study

This study will explore contextual factors that influence teacher sensemaking. This study will not evaluate or compare middle school to high school teachers, nor will it examine the outcomes of their respective grading practices. The purpose of this study will be to examine how middle school and high school language arts teachers understand and describe their respective writing grading practices. This study will examine their descriptions with the intention of understanding teachers' writing grading practices, to examine if and how pedagogy and curriculum influence the writing grading process, and to explore how teachers' sensemaking about their grading practices is affected by contextual factors, such as grade-level, classroom context, or individual teachers' beliefs. Sensemaking theory (Weick et al., 2005) will be used as a lens for viewing teachers' grading decisions, in light of contextual factors. Though other studies (Sun & Cheng, 2013; Yesbeck, 2011) have explored language arts teachers' perceptions of their grading practices, this study is unique because of the specific focus on understanding language arts teachers' sensemaking in middle school and high school, with regard to writing. This research will be guided by the following research questions:

1. What contextual factors inform middle school and high school language arts teachers' grading/ assessing practices for writing?

2. What are teachers' values and beliefs about the age group they teach, and learning and motivation in general; and, how do these beliefs influence their writing grading practices (adapted from McMillan, 2019)?
3. How do middle school and high school language arts teachers make sense of their writing grading practices?
4. What are teachers' experiences with grading writing in their respective classrooms?

Literature Overview

In the literature review section, several compelling research studies that investigate teachers' perception with regard to grading will be discussed (Bailey, 2012; Guskey, 2002; Guskey, 2009; Hay & Macdonald, 2008; Kunnath, 2016; Liu, 2007; Liu, 2008; McMillan, 2001; Randall & Englehard, 2009; Sun & Cheng, 2013; Webster, 2011; Wiley, 2011; Yesbeck, 2011). Among the existing research on this topic, this study is unique in that it will examine middle and high school language arts teachers' sensemaking with regard to their writing grading practices. As the literature review will show, grades are supposed to communicate academic performance, yet they are often adulterated by a variety of other achievement and non-achievement based factors (Brookhart, 1991). Teachers show conflicted beliefs regarding separating academic performance from students' non-achievement characteristics, such as effort (Barnes, 1985). The theme of teachers' conflicted beliefs appeared also in the form of tension between teachers' beliefs and values and external factors (McMillan, 2003). Another theme addressed in the literature was the vertical misalignment of grading practices. Teachers of different grade levels express different perceptions and different grading

practices. Guskey (2009), for example, noted significant differences between elementary and secondary teachers' grading practices and their perceptions on the purpose of grading. This study must allow for the idea that these differences are not inherently bad, as students are on a developmental continuum with different needs at each level. It is possible that grading schemas adapt to serve those developmental needs. Again, it should be noted that this study will not focus on outcomes, nor will it contribute a value judgment on any differences in practices that may emerge in the study's findings. To get to the heart of understanding how teachers perceive their writing grading practices, sensemaking (Weick et al., 2005) will be used as a theoretical framework for this study. This study will examine the contextual factors that influence teachers' sensemaking with regard to their grading practices.

With this theoretical framework in mind, this study will examine teachers' unique motivations for making writing grading decisions, at their specific grade level. As will be discussed in the literature review, for example, at the secondary level, there is a greater emphasis on personal responsibility (Queen, 2002), and consistent with this observation, Messick's (2002) respondents—including teachers, parents, and students—consistently emphasized the importance of feedback to students over parents as the purpose of grading (p. 5). Liu (2007) did not show a significant difference in secondary teachers' perceptions of grading practices with regard to academic enablers, yet the findings of his study show that middle and high school teachers had different perceptions of classroom behavior as a factor in grading, and that high school teachers were more likely to consider attendance or participation as a factor in grades. Differences in how teachers at varying grade levels perceive their grading practices are reflected in these instances. Grades serve

a variety of purposes: to motivate students, to communicate with parents, to classify students, and to measure the effectiveness of one's teaching practice (Barnes, 1985), yet teachers may very well place emphasis on different aspects therein. With the understanding that teachers' sensemaking with regard to grading may be nuanced, this study seeks to explore the contextual factors that affect teachers' sensemaking.

Rationale and Significance

This study will contribute to the existing literature on this topic by examining what contextual factors influence language arts teachers' grading practices for writing in middle school and high school. Understanding what factors influence teachers' grading practices will offer insight to educational leaders and practitioners alike. The implications of this study may be a better understanding of how and why teachers make the grading decisions that they make at their particular grade level. In addition, this study may contribute to an identification of meaningful supports that may be added to academic transitions, with regard to potential discontinuity in teachers' grading practices. The results of this study will contribute to the small body of research available examining teachers' perceptions of their writing grading practices at the secondary level. This study may also assist in understanding how language arts teachers make sense of and approach grading and assessing students' written work. Finally, the results of this study will potentially offer avenues for future research on grading.

Theoretical Framework

As this research is exploratory in nature, it follows an inductive mode of inquiry, seeking to generate theories rather than to test a hypothesis. Sensemaking will be used as a theoretical framework. This framework will allow me to explore the ways in which

teachers express their sensemaking processes, in terms of individual cognition and collective reasoning. Weick, Sutcliffe, and Obstfeld (2005) insist that when one seeks to examine an action, they must understand that what they are actually examining is the interpretation processes of the individual carrying out the action. For the sake of this research, the action—or process—being examined is teachers' writing grading practices, and in order to cultivate an understanding of how teachers perceive their practices, I will seek to uncover their sensemaking processes. It is essential to understand that sensemaking occurs in real-time, and thus as events are unfolding, these processes often highlight what Paget (1988) refers to as the “too-lateness of human understanding” (as cited in Weick et al., 2005, p. 412). In this way, the theoretical lens will allow me to examine teachers' reflective understanding of their grading decisions.

Coburn (2001) studied how teachers collectively interpreted and implemented reading policy, and though this case study does not specifically examine reading or policy implementation, her study is relevant in that she used sensemaking theory to understand how teachers “adapted, adopted, combined, or ignored pressures about reading instruction in their professional communities, and how these deliberations have shaped classroom practice” (Coburn, 2001, 147). In the same way, this study will consider that teachers are actively shaping district or departmental policies or pedagogical practices in their classrooms. As individual operators—and within communities of practice—teachers are adapting, adopting, combining, or ignoring factors in the grade-based decisions they make.

Methodology

This dissertation explores teachers' grading practices in middle school and high school language arts using in depth interviews, documents analysis, and field notes. This study will utilize an exploratory case study design (Merriam, 2001; Rubin & Rubin, 2012) to explore teachers' sensemaking with regard to their writing grading practices and assessment decisions. The case will be a single district in central New Jersey, and the sample will be limited to general education language arts teachers, 6th-12th grade.

Summary

Ultimately, this study seeks to contribute to the extant literature on issues relevant to variation in teachers' grading practices and across grade levels. This research may inform the creation of grading policies, implementation of transitional supports, and the amelioration of transitional obstacles in the district that is the foci of this study. Though the District currently has practices in place to support students in the transition from middle school to high school, those practices may be enhanced. The implications of this study may also help in identifying or clarifying supports that may be put in place in terms of grading policy or practice more broadly.

Definition of Key Terms

For the purpose of this study, a glossary of terms that will be used throughout this dissertation is offered below.

Academic Enabler- Factors that enable academic achievement but are not valid measures of achievement; they are often erroneously considered by teachers as a factor in grading.

Academic enablers include effort, ability, improvement, participation (McMillan, 2001).

Academic Enabling – Grading with consideration of non-academic factors and academic enablers (McMillan, 2001).

Educational assessment-“the process of eliciting, gathering, and interpreting evidence of student learning to describe student learning and/ or inform educational decisions”

(Brookhart, Stiggins, McTighe, & William, 2019, p. 6)

Formative assessment-the process of systematically gathering evidence about student learning to identify their current level and adapt lessons to assist a student in reaching a desired learning outcome; intended to help students and teachers adjust and improve student achievement (Heritage, 2007, as cited in Lindner, 2017)

Grading- Assigning symbolic or numerical measures to student work or generating composite student performance indicators for report cards (Brookhart et al 2016; Guskey & Brookhart, 2019).

Measurement – A process used to quantify and assign a value to how often something has been demonstrated and to what degree (McMillan, 2008)

Summative Assessment – Assessment method that provides a final grade as an outcome of learning (Brookhart, 2009)

Chapter 2

Literature Review

Teachers' grading practices are complex and nuanced. The meaning, interpretation, and implications of grades are likewise complex and nuanced. Grades are more than symbolic representations of academic achievement. According to Muñoz and Guskey (2015), the intended purpose of a grade is to communicate student achievement with regard to learning objectives or academic criteria. In fact, grades may be representative of a variety of things beyond learning objectives or academic criteria, and because of teachers' idiosyncratic practices, it is not always inherently clear exactly what the grade is representing (McMillan, 2001; McMillan, Myran, & Workman, 2002). Teachers' subjectivity, as well as potential systematic differences rooted in grade level and students' developmental needs, may potentially create a lack of vertical coherence in grading practices.

This qualitative case study will explore middle and high school teachers' perceptions of their language arts grading practices. Middle school to high school represents a significant academic transition (Ellerbrock et al., 2015; Queen, 2002), and in the district that is the foci of this study, middle school to high school also presents a potential pedagogical shift, as the Lucy Calkins & Teachers College Reading Writing Project (TCRWP) process-based workshop curriculum *Units of Study* for reading and writing are available for K-8th grades only. This potential pedagogical shift may also contribute to a shift in grading practices; the process-oriented workshop model favors formative feedback as an inherent aspect of writing instruction (Teachers College, Columbia University, 2014). In considering the variety of factors—pedagogical and

otherwise—that may influence teachers’ grading practices, the central question of this study is: What contextual factors inform middle school and high school language arts teachers’ grading/ assessing practices for writing? What are teachers’ values and beliefs about the age group they teach, and learning and motivation in general; and, how do these beliefs influence their writing grading practices (adapted from McMillan, 2019)? How do middle school and high school language arts teachers make sense of their writing grading practices? What are teachers’ experiences with grading writing in their respective classrooms? This literature review will begin with a discussion of the literature search strategy used to identify relevant literature. From there, this literature review will attempt to contextualize writing, language arts instructional models, and grading. This literature review will also include a discussion of the theoretical framework.

Literature Search Strategy

The following strategies were used to identify literature for this review: a keyword search of the Rowan University database, a review of abstracts to find relevant research, reading of all relevant research, and finally a snowball search using the reference pages of relevant studies. Initial searches were conducted using the following keywords: language arts instruction, middle school high school grading, history of grading, variation in grading, and perceptions of grading. Initial searches were limited to full-text, peer reviewed articles in English, within the last five years. Although there have been several dissertations on the topic of grading in the past decade, much of the foundational research on this topic is outside of the initial search parameters. Additional

studies and foundational research were identified through mining relevant references and citations in other research studies.

The extant literature on teachers' grading practices and perceptions has focused on elementary teachers exclusively, elementary and secondary teachers, middle school teachers exclusively, middle school and high school teachers, and high school teachers exclusively. This study will examine middle school and high school teachers, but rather than define them collectively as secondary teachers, this study will treat them as distinct grade levels. This decision was made after considering the differences that emerged in Liu's (2008) study of middle and high school teachers. Among his findings, high school teachers were more likely to consider classroom behavior, attendance, and participation as factors in grading (Liu, 2008).

The review of empirical studies presented here was limited to studies of teachers' grading practices and perceptions. The review was further limited to studies on elementary and secondary teachers, middle school teachers exclusively, middle school and high school teachers, high school teachers exclusively, and studies therein pertaining specifically to language arts. Within the body of recent research, there are only two studies that focus on the perceptions of language arts teachers (Sun & Cheng, 2013; Yesbeck, 2011), and none that focus exclusively on middle and high school language arts teachers' sensemaking of writing grading practices. This literature review presents several common themes that emerged in the extant studies: teachers' values and beliefs influence grading (Bailey, 2012; Barnes, 1985; Brookhart, 1991; Cross & Frary, 1999; Kunnath, 2016; Liu, 2008; McMillan & Nash, 2000; Sun & Cheng, 2013; Wiley, 2011), questions surround validity (Allen, 2005; Brookhart, 1991; Cross & Frary, 1999; Ebel &

Frisbie, 1991; Messick, 1989), and there is variation in teachers' grading practices (Brookhart, 1994; Guskey, 2009; Llosa, 2008; McMillan, 2001; Randall & Englehard, 2009; Stiggins, Frisbie, & Griswold, 1989; Webster, 2011). These themes offer a lens for any findings that emerge in this study. This literature review will contextualize language arts instruction and describe the evolution and purpose of grading.

Language Arts Writing and Instructional Models

Although this is a study on grading practices, it is relevant to contextualize writing instruction and to give an overview of the models of instruction in the language arts classroom, as pedagogy often informs all aspects of a teacher's practice, including grading. Grading practices in language arts are inherently nuanced and subjective. Yesbeck (2011) notes that studying language arts exclusively allows for perspective on "...one area of study that is rich in subjectivity in terms of grading written and oral responses in the forms of presentations, essays and term papers, contrary to concrete and precise answers found in most mathematics and science works" (p. 45). As Yesbeck (2011) points out, language arts as a discipline does not demand concrete, precise answers, and performance in a language arts class is difficult to quantify. This inherent subjectivity makes language arts an appropriate subject to study when looking at teachers' grading practices. In addition, as the workshop model of instruction is increasingly adopted in school districts across the nation, educational researchers must grapple with any disjuncture that occurs beyond (or within) the *Units of Study*. The school district that is the foci of this study uses the Teachers College Reading and Writing Project workshop curriculum *Units of Study* in K-8 language arts classrooms. The *Units of Study* have only been published for K-8th grade. This leaves room for a

potential pedagogical shift between middle school and high school instructional practices, which may affect grading practices, as well. Though this is a single case study, the potential disjuncture in pedagogy and corresponding practices may exist wherever the units of study are implemented.

Writing instruction. Writing instruction in New Jersey is informed by the New Jersey State Learning Standards (NJSLs). Yet mandated standards-based writing instruction is a fairly recent development. This discussion, therefore, situates writing instruction in an evolving world of educational policy, where the value of strong writing instruction was long overlooked. Before there was reduce, reuse, recycle, there were the three “Rs” of education: reading, writing, and arithmetic. These Rs represent the basic functional skills in literacy and numeracy. With the advent of the No Child Left Behind (NCLB) act of 2001, and the resulting accountability movement, there were increased efforts concentrated on improving students’ skills in reading and math in particular. Writing, however, did not receive the same attention in the policy environment. In fact, writing was cited as the overlooked “R” in education (Cutler & Graham, 2008; The College Board, 2003; Troia, Olinghouse, Zhang, Wilson, Stewart, Mo, & Hawkins, 2017). Cutler and Graham (2008) note that NCLB failed to emphasize writing, and thus measures to address writing instruction were largely absent from school reform movements. In a 2003 report, the National Commission on Writing in American Schools and Colleges deemed writing the “neglected ‘R’” and issued a “Writing Challenge” (p. 8) to increase or improve education policy devoted to writing, classroom time devoted to writing, assessment, and professional development. The Common Core State Standards (CCSS), created in 2009-2010, offered the promise of establishing clear standards for

writing instruction. New Jersey adopted the NJSLS in place of the CCSS in 2016, with minor changes (Clark, 2016).

In addition to understanding standards-based policy, it is also important to consider student performance indicators that influence policy. The Nation's Report card represents elementary and secondary students' assessment results from the National Assessment of Educational Progress (NAEP). The report cards consistently place the majority of the nation's eighth grade students at the "basic" level of writing. The basic level means partial mastery of fundamental writing knowledge and skills (National Center for Education Statistics, 2012; Persky, Daane, & Jin, 2003; National Center for Education Statistics, 2008). Though the call for attention to writing instruction has been put forth, there are barriers to implementing any change strategy. One such barrier is opacity with regard to what is known about writing instructional practices within America's classrooms (Cutler & Graham, 2008; Graham & Perin, 2007a; Graham & Perin, 2007c). With regard to instructional practices, what is known is that there are two predominant categories of writing instruction: traditional and writing workshop (the latter is categorized by Cutler and Graham as process approaches) (Cutler & Graham, 2008; McCarthy & Ro, 2011; Pollington, M. F., Wilcox, B., & Morrison, T. G., 2011). Though this study is concerned with writing grading and assessing practices, grading and assessing practices are inextricably linked with writing pedagogy.

Workshop and traditional models of instruction. The workshop model of instruction—pioneered by Donald Graves, Donald Murray, and Lucy Calkins, and currently associated primarily with the work of Lucy Calkins and the Teachers College Reading and Writing Project (TCRWP)—is humanistic and constructivist in its approach.

Humanistic approaches allow flexibility and student participation in decision making, and “personalness” (Beane, Lipka, & Ludewig, 1980, p. 85). The “workshop” model hinges upon a small-group or whole class “mini-lesson,” followed by independent practice. In this model, the emphasis is on the process of composition: pre-writing, drafting, and revising. According to Lindner (2017) the workshop model is a constructivist approach that relies on assessment, often formative, in developing scaffolds and designing instruction that is appropriate to student need. It is a teaching model that is cultivated in immediacy and responsiveness. The workshop model is not free from critique. Concerns with the workshop model mostly pertain to classroom management, structure, and the underlying assumption that all teachers and children alike are interested in the work of progress and growth (Pollington et al., 2011). Lensmire (1994) notes that if students do not cooperate with the workshop model of instruction, then even the best teachers cannot make a workshop model succeed (as cited in Pollington et al., 2011). Traditional models of instruction utilize resources like a whole class text and worksheets, with accompanying skills identified by teachers and taught to the whole class (McCarthy & Ro, 2011). Although couched in the now passé self-esteem movement of the 1980s, Beane, Lipka, and Ludewig (1980) note that traditional methods are “custodial,” associated with order and autocratic control.

Grading

This section will offer a definition of grades, a discussion of the type of knowledge being assessed, and a discussion of the importance of accuracy in grading. Grading means assigning symbolic or numerical measures to student work or generating composite student performance indicators for report cards (Brookhart et al 2016; Guskey

& Brookhart, 2019). Grades may take the form of any signifier that represents varying levels of performance, including letters, numbers, or figures (Guskey & Brookhart, 2019). Grading is characterized by indicating both a student's level of performance and a teacher's valuation of that performance (AFT, NCME, & NEA, 1990). As teachers' valuation is inevitably a factor in grading, the accuracy of their valuation must be examined. Accuracy is especially important in the current litigious climate we live in; schools that lack grading policies may be vulnerable to legal challenges when it comes to the idea of fair grading. Chartier (2003) points out that individual teacher's subjectivity on an individual assignment may not be relevant in court, but a policy supporting grade reduction not related to academic performance would be.

Assessment. Assessment is the process of gathering information to make educational decisions, offer feedback to the student regarding his or her progress, strengths, and weaknesses, make decisions about instructional effectiveness and curriculum, and to inform policy (AFT, NCME, & NEA, 1990). For the purpose of this study, the distinction between assessment and grading must be clarified: grading is intended as an indication of student learning while assessment is intended to inform teaching and learning (Carnegie Mellon University, 2019). Summative assessment evaluates student performance at the end of a learning unit for communication via grading (Brookhart et al., 2019), yet there are many formative, ungraded assessments used to identify patterns of learning and improve teaching. According to Brookhart et al. (2019), assessment can fall into a variety of categories based on the time between instruction and assessment. "Short cycle formative assessment," for example, may occur daily; "medium cycle formative assessment" may occur between instructional units;

“long cycle formative assessment” may occur at interims as a benchmark assessment (p. 11). There are two key aspects to these various types of assessment. First, there should be time left to address students learning needs before reporting (i.e. grading) (Ruiz-Primo & Brookhart, 2018, as cited in Brookhart et al., 2019), and second the assessment must be part of a broader system, with fluid communication to all stakeholders (Brookhart et al., 2019). Brookhart, Stiggins, McTighe, and William (2019) insist that formative assessment is an essential component to strong instruction yet implementing meaningful formative assessment systems remains a weakness in many districts.

A history of grading. Guskey and Bailey (2001) note that the practice of teaching and learning has used assessment for ages, dating back to the ancient Greeks, yet the nature of assessment at that time was formative rather than evaluative. In the United States, prior to the 1850s, formal education was primarily administered in one-room schoolhouses, with mixed age and background grouping of students, where progress was reported orally to parents (Guskey & Bailey, 2001). By the late 1800s, the narrative report card evolved to aid student growth and mastery. The early narrative report card consisted of a list of the skills that students had mastered and those in which they needed growth (Guskey & Bailey, 2001). Narrative report cards require knowledge of the students, as well as a significant amount of time on the part of the teacher, as they must be thought-out and composed.

The advent and expansion of public schooling affected grading practices. Massachusetts was the first state in US to mandate compulsory schooling, in 1852, marking the inception of what would become the proliferation of public schools in the United States. Between 1870 and 1910, public high schools increased from 500 to

10,000 (Gutek, 1986 as cited in Guskey & Bailey, 2001). According to Kirschenbaum, Simon, and Napier (1971), the proliferation of public schools during the late 1800s marks the advent of present grading and reporting systems in the US (as cited in Guskey & Bailey, 2001). The increasing scope, organization, and complexity of the public school system led to teacher narratives being supplanted by more formal grading systems. As with other institutions, the scope and scale of the emerging school system required more efficient and less time consuming practices. As grading practices evolved to become more efficient, their accuracy and clarity around what they were actually measuring was called into question.

A brief taxonomy of grading schemas. Brookhart (2004) identified three ways of determining grades: criterion or standards-based, norm-referenced, or student self-referenced. To highlight the distinction between norm and standards-based grades, that the latter relies on comparison to others in the class, while the former indicates that a comparison is being made to an absolute standard. That is to say, in standards-based grading, students are seeking to attain specific objectives. Frisbie and Waltman (1992) caution that teacher who employ criterion-referenced grading must describe the specific criterion to students. Though criterion-referenced grades are generally considered best practice, norm-referenced grades are often used (McMillan, 2001). Self-referenced grading, which is infrequently used, allows students to compare their current performance to previous performance in assigning a grade (Brookhart, 2004). Muñoz and Guskey (2015) point out that educators use three primary criteria for assigning grades: process, product, and progress. Product-driven criteria are favored by teachers who seek to communicate academic evaluations (O'Connor, 2002, as cited in Muñoz and Guskey,

2015). As suggested in the name, these criteria are usually based on some type of final product or summative assessment, whether in the form of a test or project. Teachers who employ process criteria, on the other hand, may not agree that product grades accurately depict learning, thus process grades account for effort and other academic enablers, or quizzes and formative assessments (Muñoz & Guskey, 2015). Finally, progress criteria accounts for growth over time, more commonly used with individualized instruction or special education (Muñoz & Guskey, 2015). What is important here is that each type of grading criteria seems to be aligned with a specific value system.

Teachers' values and beliefs. Teachers' values and beliefs arose consistently as a factor in grading. Teachers often have conflicted beliefs about separating academic performance from student characteristics, such as effort (Barnes, 1985). McMillan and Nash's (2000) study of classroom teachers in Virginia identified six themes that highlight how teachers' beliefs influence their grading. The themes were "teacher beliefs and values, classroom realities, external factors, teacher decision making rationale, and assessment and grading practices" (McMillan & Nash, 2000 p. 6). Overall, tension existed between teachers' internal beliefs and values and the classroom realities and other external factors (McMillan & Nash, 2000; McMillan, 2003). This suggests that teachers "espoused beliefs," what they purportedly believe in, do not always align with their "theories in action," what they actually do in practice (Argyris & Schön, 1974). In Wiley's (2011) study, 49% of the district's teachers who responded to the survey indicated that ability should be ignored in assigning students' grades. Yet, in responding to their actual practices, 72% of those same teachers reported that, in practice, they raise the grades of low ability students. This conflict supports the idea that teachers' espoused

theories do not always match up with their theories in use. One question posed in Wiley's (2011) research examines whether or not there is congruency in teachers grading beliefs and practices. The findings showed incongruency; for example, teachers were achievement oriented, for the most part, yet non-achievement factors, like homework and class participation, were often accounted for in grades (Wiley, 2011). The fact that teachers use achievement-related factors for the most part, but also hodgepodge factors, is supported by the literature (Brookhart, 1994; Cross & Frary, 1999; McMillan, 2001; McMillan & Nash, 2000; Wiley, 2011). McMillan (2003) isolated the factors that influence teachers' grading rationale, and ultimately, their assessment practices, including: "teacher knowledge" (including their beliefs, expectations, and values), "external factors" (including state and district policies, testing, parents), and "classroom realities" (including social promotion, absenteeism, and behavior) (p. 36). McMillan's (2003) resulting framework elucidates the existing discrepancies between measurement experts' advice on grading best practices and teachers actual experiences in the classroom.

Exemplifying how teachers' values may conflict with best practice, teachers in several studies reported inflating grades of low ability students who put forth effort (Cross & Frary, 1999; Kunnath, 2016; McMillan, 2001; Wiley, 2011). Ability and effort represent what McMillan (2001) refers to as "academic enablers" (such as effort, ability, improvement, participation) (p. 28). Academic enablers frequently influenced teachers' grading practices. McMillan's (2001) study examined potential relationships between teachers' practices and grade level, subject matter, and student ability level. In assessing how teachers used value judgments in assigning grades, he found that low-ability

students with high effort would be given a “break” with regard to grade, while the same was not true for an above-average student working below ability level (p. 21). Similar findings emerged in Cross and Frary’s (1999) study of survey results of 465 teachers—comprised of 226 middle school and 239 high school teachers—where 72% reported that they raised the grade of low-ability students. Cross and Frary’s (1999) survey results were stratified based upon the students’ preparatory track, and results showed that only 61% of college preparatory teachers reported increasing students’ grades. Similar results emerged in Wiley’s (2011) assessment of how teachers grade the “underdog” (low-ability, low-achieving student). One of the themes that emerged in Wiley’s study was that most teachers established classroom grading policies that allow for a student’s grade to rebound after a low assessment. Effort, exclusive of ability, also proved influential. Of the 122 middle and high school teachers who responded to his “Teachers’ Perceptions of Grading Practices (TPGP) in the US, Liu (2007) noted that most teachers in the US reported that they consider students’ effort when assigning grades, indicating that they reported higher grades to students with greater effort and that they were more likely to give a failing student a passing grade if that student put forth effort.

While teachers’ value systems may not be aligned with best practice, this is not the only way a teacher’s values and experience may influence his or her grading practice. Two themes that emerged in Sun and Cheng’s (2013) study of Chinese secondary school English language teachers with regard to teachers’ values in grading, were fairness and perception of what would be beneficial to students (Sun & Cheng, 2013). Similarly, Kunnath (2016) shows that teachers were influenced by their desire for students’ success and their teaching philosophies. More specifically, in his study of over 250 teachers in a

diverse high school district in California, Kunnath's results showed that a variety of grading practices exist across all demographics, but that student poverty level was an influential factor in teachers' grading practices. The quantitative results of his study showed that across all economic levels, teachers were influenced by their desire for students' success and their teaching philosophies (Kunnath, 2016). This is consistent with the findings of McMillan and Nash (2000). Among the most frequently used subjective practices, Kunnath (2016) identifies student ability level, student effort, and the inclusion of zeros; the former two were most commonly used for equity and in making determinations in borderline grades and the latter used more frequently in higher level classes.

Though many studies examine misalignment between teachers' purported values and their reported practices, other studies seek to explore what informs teachers' beliefs and values. Bonner and Chen (2019) attempt to understand the variety of factors in teachers' grading practices through considering the potential influence of learning theories. They examine existing literature about teachers' grading practices through the lens of the following learning theories: knowledge acquisition, constructivism, and behaviorism. Each learning theory has a different implication. Knowledge acquisition theory posits that academic achievement alone should be considered in grades. Constructivism allows for non-academic factors to be considered in assessing student performance ("learning is a process, not an outcome") (Bonner & Chen, 2019, p. 71), while behaviorism is primarily concerned with influencing behavior. The idea that teachers practices are influence by their values and beliefs is supported by the literature, yet sometimes, teachers values may conflict with best practices.

Grades are intended to communicate academic achievement, yet as these studies show, there are other factors that influence grading. With regard to identifying what factors make up grading, one theme that emerges in studies on grading is what Bailey (2012) identifies as a “pattern of differences” (p. 23). There appear to be a garden-variety of factors that influence that ways in which grading is approached. These factors typically fall into two categories: achievement and non-achievement based. The former is preferable for the intended purpose of grades, yet strictly academic grades are rarely reported. In reality, Brookhart’s (1991) oft quoted label of a “hodgepodge” system of grading still rings true. What is often reported is a hodgepodge of many factors, including both academic achievement and academic enablers (McMillan, 2001). With his analysis of 1,483 questionnaires completed by middle and high school teachers, McMillan’s (2001) study organizes the “hodgepodge of factors” used by teachers into four distinct categories: academic achievement, academic enablers (effort, ability, improvement, participation), external benchmarks, and extra credit (p. 28). Bailey’s ‘pattern of differences’ and McMillan’s (2001) taxonomy of hodgepodge factors point to the idea that nuances in grading can be categorized on a variety of levels.

Internalized criteria. Even when teachers use a rubric or specific criterion, their values may still affect their grading. McCallum, Gipps, McAllister, and Brown (1995) coined the term the “intuitive assessor” to describe the way that teachers showed flexible interpretation and use of standards as the basis for their grading decisions (as cited in Hay & Macdonald, 2008). The idea of intuitive grading also appeared in a study by Hay and Macdonald (2008). In this study, the researchers conducted in-depth interviews with two PE teachers in Queensland, Australia, and though their study may have limitations, their

findings, nonetheless may have implications for the ways teachers reported using their “gut feelings” and intuition in grading (p. 159). The idea that teachers may make intuitive decision while grading is an important consideration, given that in an increasingly standardized arena, variation likely still exists in teachers’ grading practices. Yesbeck’s (2011) study of middle school language arts teachers revealed that teachers’ grading practices are influenced in-part by both peers (teacher-mentors) and personal experiences. This points to the idea that grading practices are influenced by both social experiences and individual experiences. Other factors influence teachers as well. It can be inferred from other studies (Guskey, 2009; Messick, 2002; Liu, 2007) that teachers may also make intuitive grading decisions based on their communities of practice or understanding of the age group they teacher. Stiggins, Frisbie, and Griswold (1989) identify three possible reasons for grading discrepancies: differing teachers interpretation of best practice, classroom constraints, and the possibility that teachers may not know measurement theory. In investigating teachers’ sensemaking, this study will explore teachers’ experiences with grading writing, classroom contextual factors, and teachers’ values and beliefs with regard to grading students’ writing.

Horizontal and vertical misalignment. Teachers at different levels may have different perceptions and different practices when it comes to grading. According to Guskey’s (2009) survey of over 550 teachers, little variation emerged in grading practices between teachers in different subject areas, yet the differences between elementary and secondary teachers’ practices were significant. Elementary teachers viewed grades as a catalyst for parent-teacher communication while secondary teachers based their grades on their perceptions of college and career readiness. It must also be noted that each grade

level comes with a different set of expectations, and therefore teachers' values in this regard may reflect their intuitive understanding of the age group they teach or other contextual factors. In her study of ten middle school language arts teachers grading practices, using measurement theory, Yesbeck (2011) notes that only two of Guskey's (2004) six categories for grading emerged in her study. One of the categories that did not emerge as a purpose of grading was identification for special education. The fact that this was a factor in Guskey's findings and not in Yesbeck's could be attributed to context, as identification and intervention ideally occur before a student reaches middle school, and Yesbeck's study focuses exclusively on middle school language arts teachers. According to Guskey's (2009) study, teachers' perceptions of the purpose of grading also differed. Ninety percent of elementary teachers—an overwhelming majority—reported that communication with parents was the primary purpose of grades, and they were more likely than secondary teachers to carefully differentiate learning goals and standards in assigning grades. In contrast, Guskey (2009) found that secondary teachers were more likely to see grading as a crucial element of classroom management and control. Relevant contextual differences likely inform these discrepancies, including but not limited to the needs of the age group or perhaps the number of students teachers at varying levels are responsible for.

It seems that as the grade level changes, so too does the educational value system. Messick's (2002) survey of teachers, parents, and students' perspectives on grading showed a correlation between grade level and responses. As the students' grade level increased, teachers, parents, and students alike consistently reported that communicating with parents was a less important purpose of grading and providing

feedback to students was more important (Messick, 2002, p. 5). This is consistent with the increasing emphasis on personal responsibility at the secondary level (Queen, 2002); prioritizing feedback to students may be based on the increasing expectation that students are personally accountable to their own understanding of learning goals. Liu (2007) surveyed 107 teachers—52 middle and 57 high school—with a specific interest in identifying if teacher perceptions of grades varied at the secondary level. His results showed that there was no significant difference in secondary teachers' perceptions of grading practices with regard to student effort, ability, grading habits, and self-reported efficacy (Liu, 2007). Yet, middle and high school teachers did show significant difference with regard to classroom behavior (both positive and negative) as a factor in grading, moreover, high school teachers were more likely to consider attendance or participation as a factor in grades (Liu, 2007). The nuanced disparities in these studies indicate that there may be fundamental differences in the way teachers at varying grade levels perceive grades. Or, the differences may be indicative of teachers' intuitive understanding of the changing nature of the age group they teach. For example, as the emphasis on parent involvement decreases as students move through the school system, teacher's values regarding grading as a form of communication with stakeholders may adjust (Queen, 2002).

Most of the empirical research considered for this study explored or examined middle and high school teachers collectively, as secondary level teachers (Bailey, 2012; Cross & Frary, 1999; Guskey, 2002; Guskey, 2009; McMillan, 2001), while Liu's (2008) study of teachers' perceptions of their grading practices opens a door for this research, as it examines whether school level affects teachers' perception. The findings of his study,

for example, show differences in perceptions of classroom behavior with relevance to grades. It is possible that other differences exist between these two grade levels when considering teachers' sensemaking of grading and assessing students' writing.

Variation in grading practices. Teachers are faced with a variety of decisions around assigning grades. They must determine the characteristics to assess, the assessment methods, amount of data to collect, how data should be weighted for a composite index of performance, and how to use scores to track achievement (Stiggins, Frisbie, & Griswold, 1989, p. 6). Once assigned, grades serve a variety of purposes, viewed by teachers in a variety of ways: to motivate students, to communicate with parents, to classify students, and to measure the effectiveness of one's teaching practice (Barnes, 1985). Many school districts have made attempt to obviate problems with grade validity, such as imposing grading policies or guidelines. Yet, schools with guidelines and policies are no exception to the rule; variation in individual teachers' grading practices persists (Brookhart, 1994; Guskey, 2009; McMillan, 2001). Even when teachers have a scoring criterion, such as Llosa's (2008) study of the English Language Development Classroom Assessment, teachers may not apply the criteria consistently. Randall and Englehard (2009) found that most teachers primarily use grades as a measure of achievement, yet there remain teachers who consider other factors, such as ability, behavior, and effort. Their study, however, did not offer stratified data based on teachers grade level responses. According to Webster (2011), teachers reported inconsistency in several areas with regard to assigning grades: in terms of process, product, and data pertaining to student progress. Stiggins, Frisbie, and Griswold's (1989) case study of 15 secondary teachers examines teachers' practices with regard to nineteen dimensions of

recommended practice. Their findings, as echoed in other research, show a discrepancy between teachers' intentions, their practices, and what experts deem best practices in grading. Stiggins, Frisbie, and Griswold identify three possible reasons for the discrepancies that emerge: Teachers interpretations of best practice may vary, classroom constraints may exist, and finally, teachers may not know measurement theory. Each of these conditions may vary for individual teachers. Brookhart et al. (2019) insist that systemic coordination is essential; classroom teachers practices should be consistent with other teachers, principals should oversee meaningful practices, and school districts should ensure that all students receive equal protection.

Validity. Grading validity is questioned in the literature on two levels: what the grade represents and how the grade is used (Messick, 1989a, 1989b, as cited in Brookhart, 1991). According to Messick (1988), “validity is an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores and other modes of assessments” (as cited in Messick, 1989). There are other threats to the validity and reliability of grades: inconsistent criteria and inclusion of factors aside from the educational outcome being assessed (Brookhart et al., 2019). With his definition, Messick points to the intersection of assessing what a score actually means and the ethical implications of how the score is used. In addressing this first problematic aspect of grading, Brookhart (1991) made a crucial observation, which has remained a sticking point for future studies. She writes, “A hodgepodge grade of attitude, effort, and achievement, created in an attempt to provide positive feed- back to the student about himself or herself, is not the answer. Such a hodgepodge grade also falls down under a

validity check; it does not possess the characteristic of interpretability” (p. 36).

Brookhart’s “hodgepodge” assessment of grading proved a seminal critique of grades and their validity, or lack thereof. As a grade is made up of many unidentifiable aspects, it cannot be interpreted and thus loses validity as an academic measure. Ebel and Frisbie (1991) offer three factors in grading that fuel the controversy over grading practices; they are: technical challenges in measuring academic achievement, differing educational philosophies, and a teacher’s conflicting role as advocate and judge.

Validity as an academic measure is not the only problem. The literature points to a problem with how learning is assessed but also with how the assessment is communicated (Allen, 2005). Using Messick’s (1989) conceptualization on validity, Brookhart insists that teachers’ decisions with regard to grades are affected by the social implications of grades. The social implications she noted were grades being used: by some parents as incentives, for participation in athletics, for advanced placement and tracking, for participation in certain projects, or for rewards, like the valedictorian speech, in college admission, and more broadly, in students’ self-conception and motivation. Brookhart observed that some of these uses are valid, some are not. She advocates for valid grading but also for grades to be used for achievement related decisions, not for nonacademic privileges, such as participation in sports. She insists that until these important distinctions are made, teachers are likely to continue to be conflicted over grading.

Punitive grading. As noted earlier, secondary teachers may view grades as a component of classroom management and control. This section will explore what that looks like in practice. There is an increasing association in popular literature between the

“zero” grade and equity in grading. In McMillan’s (2001) study, about half of the 1,483 middle and high school teachers surveyed indicated that a zero should be assigned, regardless of resulting in a failure for the term. The other half of teachers surveyed indicated that they would lower the grade, yet not enough to warrant a failure for the term. The written responses of teachers indicated an emphasis on fairness to the students. Guskey (2004) notes that zeros are sometimes erroneously used by teachers as a means of punishment, which is not supported by research as a valid practice, and moreover, it can have deleterious effects on student motivation and learning (Guskey 2004a). Canady and Hotchkiss (1989) call zero grades “one of the most punitive and damaging weapons in a teacher’s grading arsenal” (p. 69). The authors cite that zero policies unfairly affect students with home situations that do not allow them to complete work outside of the classroom. In terms of teachers beliefs and values, teachers’ views with regard to the zero grade could be an indication of “behaviorist” inclinations, as assigning a zero is not a measure of academic performance, but rather a measure that indicates failure to comply with a specific behavior or expectation (completion of assignments). In discussing borderline scenarios where students are otherwise showing a learning progression, McMillan (2019) refers to zeros as egregiously unfair. McMillan (2019) notes, “...the zero assumes no knowledge or proficiency and unduly skews the overall score in a negative direction” (p. 105). Contrary to popular arguments against abolishing zero grades in favor of a minimum grading system, in their study of seven years’ worth of data from an urban high school in California, Carey and Carifio (2012) offer that a minimum grading system neither promoted grade inflation nor social promotion. Instead, they offer that it resulted in an economic benefit to the school with a decrease in attrition and

reduced costs with regard to summer school. While grades are sometimes used to influence students' effort or behavior (Guskey, 2009), punitive grading practices may have deleterious effects.

Best Practice

Guskey and Brookhart (2019) encapsulate a knowledge-base from the last century of research on grading into an eight chapter book, *What we know about grading: What works, what doesn't, and what's next*. Guskey and Brookhart (2019) introduce their book by pointing out a problem that is—unfortunately—not unique to research on education; many practitioners get their “expert” advice from sources that may lack a scientific approach to research. The authors point out that many educators today often turn to books and blogs for advice on grading, even though many of these authors do not include cited research-based evidence. As their respective research contributions to this topic—spanning several decades—(Brookhart, 1991; Brookhart, 1994; Brookhart, 2004; Brookhart et al., 2019; Guskey, 1994; Guskey, 2002; Guskey, 2009; Guskey & Bailey, 2001) is foundational, their recent book will serve as a key resource in examining and understanding current best practices in grading for the purpose of this research study. Guskey and Brookhart (2019) note that research in grading to date has identified several actionable factors to ensure best practice in modern grading. These factors pertaining to the following: goals, feedback, and accuracy. Given the significance of each of these factors, and their relevance to understanding grading schemas, each will be treated separately.

Start with clear goals. In order to understand exactly what is reflected in a grade, the learning goals must be clear, and they must be clearly articulated to students. Guskey and Brookhart (2019) insist that grades that are accurate and meaningful must begin with “...congruence among the major elements of the teaching and learning process: *curriculum, instruction, assessment and grading*” (p. 215). These four factors must be aligned, and they must build upon each other (Guskey & Brookhart, 2019). Furthermore, Guskey and Brookhart insist that there must be both content (the knowledge students are expected to attain) and process (what they are expected to do with that knowledge) identified within these goals. This component of grading best practice is foundational, as it marks the difference between memorizing and forming enduring understandings (Guskey & Brookhart, 2019).

Feedback. It is essential to the learning process that students know how their performance fairs with regard to expectations. In addition, they must have a guidance with regard to what their next steps should be in the learning process. According to Guskey (2017). It must be clear to students and to their guardians that grades reflect *where* a student is in terms of the learning process; they are not to be taken as a value judgment on the learner (as cited in Guskey & Brookhart, 2019). In this way, in order for grades to be used as effective feedback, they must accurately reflect academic achievement. McMillan’s (2019) analysis of extant research on teachers’ perceptions on grading shows that to a certain extent, classroom grading is individualized. He notes that there is more research on whether these variations may positively influence student motivation. He urges a discourse on this topic, and at the very least, including asking teachers to express how their values and beliefs about learning and motivation may affect

their grading practices. With regard to feedback, he insists that what is important is the meaning that students attach to grades, and for that reason, teachers must confer with individual students to clarify meaning and interpretation of grades.

Accuracy. Grading accuracy will be discussed in two topical areas: criteria and categories. First, in examining extant literature on grading criteria, Guskey and Brookhart (2019) note that two problems emerge: lack of clarity on the criteria by which grades are assigned and nuance among teachers regarding what criteria receive emphasis. Presently, clarity regarding criteria is a standard recommendation to reduce variability in grading, though this was not always the situation (Guskey & Brookhart, 2019).

Reliability in grading can be reduced to three essential components “criteria, consistency, and categories” (Guskey & Brookhart, 2019, p. 26). To that end, the authors offer several suggestions, such as: use a rubric, checklist, or point scheme; offer a model response; bracket and offer separate feedback for things that are not specified in the learning criteria; grade anonymously to obviate subjectivity; and ask a colleague to regrade or evaluate the grades assigned (Guskey & Brookhart, 2019). Finally, Guskey and Brookhart (2019) insist that grading scales that have fewer categories that articulate levels of mastery are more reliable, and they offer students better guidance to improve. Similarly, Link (2019) notes that the addition of more categories—like pluses or minuses—to a grading system reduces reliability. The final suggestion with regard to accurate grades has to do with composite grading. Guskey and Brookhart (2019) insist that rather than a single grade on a report card, students would be better served by increasing clarity around what the grade actually means by, for example, parceling grades into categories that include “product criteria” (academic achievement), “process criteria”

(academic enablers and behaviors), and “progress criteria” (to reflect student gains) (Guskey & Brookhart, 2019, p. 221).

Misalignment With Best-Practice

Cross and Frary (1999) insist that there is consensus among measurement experts that grades in academic subjects should be based on measures of current academic achievement exclusively. The authors defend this position by explaining the problematic use of non-academic factors as it exists, and then they call for the measurement community to aid secondary teachers in “appreciat[ing] the need to make a clear distinction between measured academic achievement and informal assessments of their pupils, basing grades on the former and using the latter for instructional or classroom decisions” (p. 56). They note that this distinction is especially crucial at the secondary level, where the objectivity of grades has implications for educational and career-based decisions. The distinction between elementary grades and secondary grades raises an ethical argument regarding secondary teachers’ practices, since the importance of objective grades at this level poses serious implications for the future decisions of the students.

Cross & Frary (1999) support the argument that non-academic factors—like growth, ability, effort, conduct—which often make their way into reported grades should not be considered in determining academic achievement grades. Brookhart’s (1991) contention is that grades not only include these non-academic factors, but also that the use of grades has become problematic and more rooted in social than academic factors. Though she coined the disparaging term hodgepodge grading, she also plays the apologist in explaining teachers’ intentions in including nonacademic factors. Nonetheless, she

concludes that the implications of hodgepodge grading are problematic to grade validity and have manifold social implications, as well. Canady and Hotchkiss (1989) insist that if teachers assign value to behavior, participation, respect for teacher, and punctuality, as examples of non-academic factors that make their way into grades, then they should report a students' performance in those areas separately.

Hopkins (1998) insists that using one's natural ability as a factor in grading for achievement is a practice that can neither be maintained nor defended as sound, regardless of its popularity in practice. According to measurement experts, it is also unreliable to use growth or student improvement in determining grades because both are unreliable (Ebel & Frisbie, 1991; Gronlund, 1985; Hopkins, 1998). Cross and Frary (1999) insist that recommended practice also encourages teachers not to consider effort when determining grades, yet in their study of over 4,000 teachers, 25% reported that they inflated grades based on students' efforts "fairly often," similarly, 39% of teachers reported considering conduct and attitude (p. 61). It is relevant to refer to a discrepancy between teachers' "theories-in-use" (what they do in practice) and their "espoused theories" (what they purportedly believe) (Argyris & Schön, 1974) here, as 81% of the teachers surveyed reported that they believed that achievement, effort, and conduct should be reported independently.

It is clear that studies of practice reveal that teachers are not following the best practices put forth by measurement experts. This could be attributed to a lack of accurate messaging, resistance to change, or lack of awareness. Cross and Frary (1999) support the latter as they note that parents, administrators, and teachers may not be considering the negative implications of their "hodgepodge" grading practices (p. 70). Researchers

(Brookhart, 1991; Cross & Frary, 1999) have put a call out to the measurement experts to get the word to teachers on what *should* be done. Yet, there are complications to getting the word out. As Cross and Frary (1999) point out, measurement advice is sometimes esoteric and opaque nature, and it is not necessarily the reading that is consumed by professionals in the field: “Although measurement specialist discuss these issues in textbooks and technical journals, others offer advice that is often at odds with recommended practice in journals widely read by teachers and administrators” (p. 70). To highlight this point, Cross and Frary (1999) use the example of Orenstein’s (1994) article published in a journal put forth by the National Association of Secondary School Principals, which offered a succinct list of ten guidelines on grading that included recommendations (Ornstein, 1994). Though Orenstein, a Professor of Education and Social Policy at St. John’s University, correctly notes that the purpose of grades is different for teachers at different levels, he posits advice that is contrary to what experts suggest.

Cross and Frary (1999) use his work as an example of misleading, yet widely-read, advice. Among his recommendations are leaving a modicum of discretion to teachers, but he also insists that grades include consideration of progress, “attitude as well as achievement, especially at the middle level,” and he insists that while grades should be comprised of mostly “objective sources” that “... some subjective sources should also be considered” (Ornstein, 1994, p. 63). Cross and Frary (1999) note that his advice conflicts with the existing—albeit less accessible—literature on the subject. This highlights a problem: what passes as popular wisdom on this topic is not necessarily in-line with what is supported by the experts.

Theoretical Framework

The theoretical framework for this paper is sensemaking theory, which reflects a theme that emerged in the extant literature on teachers' perceptions of grading: a variety of factors influence teachers' grading decisions. Sensemaking theory will be applied to understand teachers' perceptions of their routinized grading processes and grading precedents within their communities of practice. In this way, sensemaking will be applied to understand how social and institutional contexts influence teachers' grading practices. Sensemaking, a process for organizing thought, is a mediator between actors and their contexts. As such, there are activities in sensemaking, such as "noticing and bracketing" (where one interprets something that has occurred) and labeling and categorizing that help individuals construct their understanding of the world in which they operate (Weick et al., 2005). With regard to labeling and categorizing, categories have flexibility as they are social constructs, fit to local contexts (Weick et al., 2005). Brown, Colville, and Pye (2014) insist that people are reflexive participants in situations, and that beyond individual interpretation, sensemaking involves actively shaping situations. They write, "People engage in partially overlapping processes in which they construct 'realities' and then retrospectively make sense of them in continuing dialogue of discovery and invention in which identities and social worlds are concomitantly referenced and fabricated" (p. 267). This explanation of sensemaking acknowledges the idea that people actively engage in understanding processes as they occur, as well as retrospectively.

Coburn (2001) argues that sensemaking is a social process and that it is rooted deeply in the contexts in which teachers operate. Colburn (2001) identifies three

subprocesses that are characteristic of collective sensemaking. First, she observed that teachers constructed understanding through social interaction; that is, when they encountered new messages about implementing reading instruction, they relied on collegial interactions to interpret and understand the messages. Second, she observed “gatekeeping,” where teachers determined whether to implement or dismiss the message, as they understood it, and expressed their reasoning to colleagues (Coburn, 2001, p. 154). Finally, Coburn (2001) observed that teachers collectively worked out technical and practical details. With that said, this research will attempt to explore the three processes in sensemaking. Unique to this research, these three processes will account for collective and individual sensemaking by taking the following shape: taking cues from the social establishment, taking cues from the institutional establishment (understanding policy, grading frameworks and expectations); interpreting and adapting or resisting grading practices. Coburn (2001) insists that teachers make meaning out of policy collectively. In her research on how teachers made sense of reading policy and ultimately put it into practice, Coburn explains how sensemaking processes unfolded between teachers in her research: “...in conversations with their colleagues in formal and informal settings, co-constructed understandings of messages from the environment, [teachers] made decisions about which messages to pursue in their classroom, and negotiated technical and practical details of implementation” (Coburn, 2001, p. 146). Similarly, this research will examine how teachers collectively, within their given contexts, make sense of and adapt their grading practices.

One theme that emerged in the literature was that teachers at different grade levels perceived responsiveness to different student needs, which informed their grading

practices. Similarly, Coburn's (2001) research, using sensemaking to understand teachers' implementation of reading policy, shows that location mattered. That is, with whom teachers spoke—and in what particular setting—affected how they made sense of a consistent message. The idea that context contributes to the sensemaking process will be considered as this research examines multiple communities of practice and differing grade levels, as well. Coburn (2004) identifies sensemaking as the facilitator between the broader environment and classroom practice, in so far as teachers' understanding of things as they are, how they should be, and how things are actually done in practice is socially constructed. Teachers' pedagogical beliefs and the curriculum at each grade level will be considered as a feature of location, or a contributor to each teachers' unique context. As it pertains to grading practices, this theoretical framework will help elucidate the way that teachers perceptions are formed and mitigated by the different communities in which they practice.

In addition to using sensemaking theory as a catalyst for exploring context, this theoretical framework will also be used to explore what motivates nuances in teachers' grading practices. Since nuances in grading practices was another important theme that emerged from the literature, this theoretical framework is an appropriate theoretical lens through which to explore this phenomenon. Practitioners act “thinkingly,” meaning they simultaneously interpret knowledge with established frameworks (Weick et al., 2005, p. 412). Teachers' grading practices represent an established framework operating within a broader context. That is to say, teachers in the same community of practice have access to the same institutional information on grading, yet teachers individual and collective grading practices may still operate with nuance. Sensemaking theory focusses on the

organization of thought within a framework, so it is an appropriate lens through which to understand any nuances that emerge. Weick (2005) notes that sensemaking is driven by plausibility rather than accurate perceptions, and plausibility simply means how different people use the same information based on their context. For example, a teacher will apply sensemaking to grading based on his or her lens as an educator, while a child will make sense of a grade through his or her lens as a student, and so forth for parents, and administrators. This case study will examine teachers' sensemaking with the understanding that an individual's sensemaking is not necessarily rooted in accuracy.

Summary

This study focuses on middle and high school language arts teachers exclusively for several reasons. First, language arts grading is inherently subjective, since responses are often based on individual interpretation. Second, there is a potential disjuncture between the process-based workshop model and more traditional high school instruction. Overall, across decades of research, there has been little observable change to teachers' grading practices. Hodgepodge grading has remained a theme in the literature since the term was first coined by Brookhart in 1991. It seems that one enduring understanding in the studies on grading are that teachers persist in "hodgepodge" grading practices. One gap in the literature is that there is a dearth of studies related to specific disciplines, such as language arts, and no existing studies that specifically examine the teachers' sensemaking of their grading and assessing of student writing in middle school and high school language arts.

Chapter 3

Methodology

Any study that deals with unpacking the way people make sense of their lives, in any regard, requires sensitivity and careful planning. At its most basic level, this study seeks to understand the sensemaking processes of teachers. The design of a research study is crucial, as the research design determines how the study will use investigative techniques (Krathwohl & Smith, 2005), and the design of this study was carefully planned with that in mind. This chapter outlines the study's research design, including strategies for conducting interviews and collecting and analyzing material culture. This chapter will present a detailed explanation of the case study design and why it is an appropriate methodology for this study. This chapter will begin with a description of the problem: potential lack of coherence in teachers' grading practices; the purpose of the study and research questions; the significance and role of the researcher; the design of the study; the setting and participants; data collection; and analysis. This chapter will conclude with a discussion of ethical issues, the Institutional Review Board process, and a summary.

Research Questions

In designing this case study and reviewing the literature, several questions emerged to guide the exploration of teachers' grading practices. These questions focus on themes from the extant literature; they will alternately explore teachers' internal beliefs and external factors.

1. What contextual factors inform middle school and high school language arts teachers' grading/ assessing practices for writing?

2. What are teachers' values and beliefs about the age group they teach, and learning and motivation in general; and, how do these beliefs influence their writing grading practices (adapted from McMillan, 2019)?
3. How do middle school and high school language arts teachers make sense of their writing grading practices?
4. What are teachers' experiences with grading writing in their respective classrooms?

Research Design

This study used an exploratory case study design (Merriam, 2001; Rubin & Rubin, 2012) to understand and analyze teachers' cognitive understanding of their own grading practices and decisions in assessing students' writing. Not only is this study examining teachers' explanations of their writing grading practices, it is also inviting teachers' reflections on their actions. The latter will get at the heart of their sensemaking processes and the variables that influence their sensemaking around grading. Merriam (2001) suggests that case studies can be defined by their unique features; they are "particularistic, descriptive, and heuristic" (p. 29). Merriam (2001) defines particularistic in terms of focusing on something specific (a case, event, or program) and examining what the case may reveal about a phenomenon. Merriam defines descriptive in terms of what the study produces—the result of a case study is a thorough description of the case or phenomenon being studied. As such, case study is an appropriate design for a study examining teachers' sensemaking with regard to grading, as ultimately, this research seeks to offer rich descriptions of teachers' sensemaking, and the existence and interplay of contextual factors that influence teachers' thought processes. Finally, Merriam defines

heuristic in that case studies are exploratory and as such, the study and the researcher are pursuing the discovery of meaning. As a case study, this research is intended to highlight the complexities of grading, include rich description of teachers' perception of educational context, and specific grading processes at the individual level, and within a larger educational system. The philosophical orientation of this research is interpretative, as the research seeks to understand a process and considers that the experience of each individual engaging in that process will be nuanced, as "[m]ultiple realities are constructed socially by individuals" (Merriam, 2001, p. 4; Rubin & Rubin, 2012). This research follows an inductive mode of inquiry, which means that it seeks to generate theories rather than test or prove an existing hypothesis about an educational phenomenon (Merriam, 2001; Rubin & Rubin, 2012). This is an appropriate approach for this study because although theories about grading exist, those theories may not adequately explain the phenomenon of grading of students' writing in middle school and high school.

Setting

There are boundaries in case study design that establish specific limits to the phenomenon being studied (Merriam, 2001, p. 27). Though this study explores middle and high school language arts teachers' sensemaking of grading writing, it remains a single case since these grades are bound by the same department, in a single school district. In discussing the setting of this case study, it is also important to understand the boundaries that are generated by single case study design. This study follows a single case design for several reasons. First, in order to understand the grading processes of individual teachers, this study needed to eliminate the possibility that there are differing

district policies around grading, therefore this case study will examine a single school district in New Jersey. All of the teachers in this study work in the same district and are all under the same district policies with regard to grading. This study is also bound to a single subject—language arts—and to teachers of sixth through twelfth grades.

The school district. According to The State of New Jersey Department of Education Website “NJ School Performance Report” for the 2017/ 2018 school year, this district serves approximately 9,500 students in grades Pre-K through twelfth grade. In 2017/2018, this district had a fairly equal distribution of male and female students, with approximately 11% economically disadvantaged, and approximately 20% students with disabilities. It should be noted that these trends vary between the three middle schools and two high schools that will be the foci of the study, with some schools receiving Title 1 funds and others not. This district’s racial and ethnic distribution is predominantly white, at over 85%; 7% of the student population is Hispanic; less than 2% black or African American; less than 3% Asian, Native Hawaiian, or Pacific Islander; less than 1% English Language learners; and less than 2% two or more races. The dropout rate for the district in 2017/2018 was .3%, lower than the state average of 1.2%. Of the three middle schools, the student populations ranged from 650 to over 1,000, according to The State of New Jersey Department of Education data from the 2017/ 2018 school year

Sample and participant selection. I selected participants for this study using stratified purposeful sampling. The sampling is purposeful in that all interviewees share a specific attribute (Rossman & Rallis, 2017); the attribute they share is that they are all middle school language arts teachers in a specific district. Under the umbrella of language arts teachers, there are general education teachers, in-class support (ICS) or in-

class resource (ICR) teachers, resource room teachers, and language learning disabilities (LLD) teachers. This study will examine general education teachers, with the understanding that most general education teachers are also responsible for additional levels of instruction (that is, although a teacher may be a general education teacher, they may also have honors and ICS or ICR instructional groups). Patton (2002) defines stratified purposeful samples as “samples within samples” (p. 240). The purpose of stratified purposeful sampling is to identify variation rather than a common core, though that is not to say that a common thematic core won’t emerge in the data (Patton, 2002). Stratified purposeful sampling was an appropriate sampling method for this study because the overarching interest is in what informs teachers’ perspectives.

I began the sampling process by identifying prospective participants in three middle schools and two high schools in this school district. Each potential participant fit the criteria of the purposeful sampling design: they were general education, language arts teacher of grades 6-12, in the same school district. There were approximately 45 potential participants, and each was contacted via email with a description of the study and a request for voluntary participation. I received a total of nineteen positive responses from seven middle school teachers and twelve high school teachers, and I subsequently scheduled and conducted interviews with each of these teachers.

Data Collection

All but two of the nineteen interviews that were conducted for this study took place in the participants’ classrooms. One study was conducted in the interviewer’s classroom and the other in a neutral location, as no other private locations were available in these instances. Two respondents requested to be interviewed with other teachers. I

interviewed those respective teacher pairs together at their request. In one instance, the teachers taught the same grade and were highly collaborative with similar responses, and in the other instance, the teachers were philosophically aligned yet taught different grades and thus had more variation in their respective responses. Participants for this study were only interviewed once, though participants were offered the opportunity to follow-up with the researcher for any purpose; only one participant contacted me to clarify her practice in terms of conferring with students. Overall, teachers' interview responses were thoughtful and reflective, and many teachers—having been sent the questions in advance—prepared notes for the interview. Each interview was recorded using Rev, an app for digital voice recording and transcription. The audio file and transcription were then assigned a numerical code by the researcher. After each interview, I recorded field notes about the exchange, the setting, or the conversation, along with analytic memos to track emerging ideas about the data.

Interviews. Data collection for this study included nineteen person-to-person teacher interviews (Rubin & Rubin, 2012). I created and used an interview protocol to guide the interviewing process (see Appendix D). The nature of qualitative interviews is generally “more open-ended and less structured” (Merriam, 2001, p. 74). Consistent with this type of interview, the flexibility allowed me to be responsive to the respondent, the respondent's worldview and to any new ideas that were introduced into the discussion (Merriam, 2001). The interviews were recorded, using the Rev Voice Recorder app, and transcribed following each interview. Each transcription was assigned a unique identifier and was organized according to a number and letter assigned to each respondent, with a master list maintained on a private, password protected server.

Documents. Another form of data collection that this study used was the gathering of documents. Merriam (2001) uses the term “documents” to refer to written, visual, and tangible material that bears relevance to the study being conducted. Each document collected was situated in the context of the individual teacher and the grade level, rather than studied in isolation. Hodder (2012) notes that “[t]he text can ‘say’ many different things in different contexts...the written text is an artifact capable of transmission, manipulation, and alteration, used and discarded, reused and recycled...” (p. 112). The documents gathered for this research consisted of administrative meeting notes, collaborative departmental notes pertaining to grading concerns (created by one middle school language arts department), and supplement grading materials provided by individual teachers. In establishing the process for document analysis, I followed Altheide’s (2011) six steps for document collection and analysis: 1. I selected a problem to investigate 2. I familiarized myself with the context 3. I reviewed examples (e.g. I reviewed my own teaching materials to confirm where grading policies are typically outlined) 4. I listed categories to guide collection 5. I practiced the protocol 6. I revised and refined the protocol. To that end, each document was analyzed using the document protocol (see Appendix D). The document protocol consists of four questions, which alternately explore the nature of the document, the context in which it was created, and what is explicitly and implicitly communicated by the document.

Field notes. Rossman and Rallis (2017) encourage researchers to record their observations in the field in the form of field notes: “...the written record of your perceptions in the field” (p. 172). The two primary components of field notes are descriptive data of the researcher’s observations as well as the researcher’s comments on

the data or the research itself (a “running record”) (Rossman & Rallis, 2017, p. 172).

Field notes consisted of my own notes before, after and during each interview. These notes described the rapport of the interview, the classroom and school environment, and my general perceptions.

Research Validity

Assessing validity begins with looking at how the study was conceptualized, and then examining the processes for how data was collected, analyzed, interpreted, and finally, how the findings are presented. The design of the research study and the researcher’s ethical conduct are at the forefront of Merriam’s (2001) presentation of a valid and reliable qualitative study. I will ensure validity through pilot tests and data triangulation. According to Yin (2009), one of the major strengths of data collection for case studies is the potential, or need, for varied data sources. Yin insists that the use of varied data sources allows for a broader range of issues—historical and behavioral—to be addressed. He also insists that these “converging lines of inquiry” encourage triangulation of the data (p. 115). To ensure that there are converging lines of inquiry in this study, interview transcripts, documents, and field notes will be analyzed and triangulated.

As this study will examine teachers’ sensemaking, the results of this study will rely on my interpretation. Long-term observation, “repeated observations of the same phenomenon” (Merriam, 2001, p. 204), will be used through conducting interviews until achieving saturation (Rubin & Rubin, 2012). I also sought participatory research through piloting the protocol questions. In the fall of 2019, the interview questions were piloted to a literacy coach and a district interventionist, both former language arts teachers. The

purpose of these pilot tests was to garner feedback on the interview process and the clarity of the interview questions. The pilot test was also used to ensure that the data yielded was appropriate to the research questions at the heart of this study. Feedback, and a review of the data yielded from the pilot interviews, was used to revise the interview protocol.

Credibility. Patton (1990) insists that qualitative research is context-dependent, and as such, it should be taken as perspective rather than a generalizable truth (as cited in Merriam, 2001). I will take measures to ensure that this research has internal validity—credibility. Toma (2006) defines credibility as the extent to which the participants agree with the accuracy of the researcher’s interpretations and descriptions. The key to ensuring credibility, according to Toma, is bounding the case within its context and using rich description. Miles and Huberman (2004) insist that the researcher is an instrument in the research, and their credibility relies upon their familiarity with the phenomenon and setting being examined, interest in the concepts, a multidisciplinary approach, and the ability to elicit information from participants (as cited in Toma, 2006).

Reliability. Reliability in terms of replicating findings is not necessarily possible within a qualitative case study; Merriam notes: “Because what is being studied in education is assumed to be in flux, multifaceted, and highly contextual ...achieving reliability in the traditional sense is not only fanciful but impossible” (p. 206). For these reasons, Lincoln and Guba (1985, p. 288) encourage qualitative researchers to focus on “dependability” or “consistency” (as cited in Merriam, 2001, p. 206). To that end, this research focused on three aspects of obtaining dependable, consistent analysis of the data, concerning the investigator’s role, triangulation of the data, and ensuring an audit trail

(Merriam, 2001). To ensure dependability and consistency with regard to the data, I will triangulate the data as mentioned previously. To ensure dependability and consistency with regard to protecting the data collected, I will offer thorough description of how the data was collected, how the categories were generated, and I will offer clear and thorough insight to the decision-making process (Merriam, 2001).

Data Analysis

Merriam (2001) points out the researcher in a single-case study will face challenges making sense of the data, given the potential of disparate information being presented in the varying data sources. For this reason, data management and organizational strategies are crucial (Merriam, 2001). Rossman and Rallis (2017) define analysis as “the process of bringing order, structure, and meaning to the mass of collected data” (p. 237). They offer the following eight phases of analysis: “1. Organizing the data, 2. Familiarizing yourself with the data, 3. Identifying categories, 4. Coding the data, 5. Generating themes, 6. Interpreting, 7. Searching for alternative understandings, 8. Writing the report” (Rossman & Rallis, 2017, p. 237). Analysis is iterative, so these phases will not be linear, as I will likely move between and among these phases (Rossman & Rallis, 2017). These authors caution that the analysis process is ongoing, and they offer the following advice, which will be followed in this study: refer often to the theoretical frame, yet remain open to new discoveries; let the questions guide the work; thoughtfully modify data gathering as needed; document all decisions for the audit trail; write, reflect, and discuss ideas; continue reading; and be creative.

The analysis process for this study began with familiarizing myself with the data. I oriented myself with the data by reading all of the interview transcripts, as well as

reading and rereading my field notes. After reading all of the transcripts, I uploaded the transcript documents to the MAXQDA Analytics Pro 2020 software program, and then I began coding chunks of text. Since the research questions of this study seek to understand teachers' perceptions, and the actions/ processes they engage in (Saldaña, 2016), the coding methods for the transcribed interviews used a hybrid approach to coding the data using Provisional, Process, and In Vivo coding methods. I developed a list of provisional codes based on categories I anticipated (Saldaña, 2016). This was an appropriate coding method since I am familiar with the roles and responsibilities of the individuals I interviewed, and provisional codes allowed me to organize anything that was stated or implied with regard to the three subprocesses of sensemaking. The initial codes were taking cues from colleagues, taking cues from students, taking cues from parents, taking cues from precedents, taking cues from district, taking cues from curriculum, taking cues from pedagogy, implementing grading, adapting to policy, modifying policy. Some provisional codes were adapted from themes that emerged in the literature review, such as: discussing practice and referencing standards; discussing practice and norm-referencing; discussing practice and students self-referencing; discussing process, product, or progress. Other provisional codes were also informed by the literature, such as Ebel and Frisbie's (1991) factors in grading: technical challenges in measuring academic achievement, differing educational philosophies, and a teacher's conflicting role as advocate and judge. It should be noted that these provisional codes were useful in approaching the data, but not all of the codes were utilized in the coding process, nor were all of them necessarily relevant to the themes that emerged.

In addition to provisional coding, process coding and In Vivo coding were used. Process coding involves attaching gerunds or descriptive phrases to identify actions described by teachers (Saldaña, 2016). Process coding was appropriate as a first cycle coding method for this study because this study is examining specific processes in which teachers engage. In addition, process coding aided in recognizing which actions occurred multiple times within an interview or across interviews. For example, sixteen of the twenty-two first cycle codes that appeared with a frequency of ten or more times were process codes. These codes allowed me to identify specific actions or processes to the respondents' descriptions. This also allowed nuance prior to theming the data, for example, discussing time constraints and discussing time both pertained to the theme of time, yet the process coding addressed nuances therein. Finally, In Vivo coding was utilized less frequently than the aforementioned, yet it was relevant in moments where teachers' exact words or phrases were relevant (Saldaña, 2016). With that said, there were moments where teachers used similar phrasing across interviews, and in these moments, In Vivo codes were used to capture teachers exact words. One example of an In Vivo code that appeared across interviews was "noting the gray area" in grading writing. This code was useful in capturing the essence of teachers' sensemaking processes, in their own words.

After the nineteen transcripts and field notes were coded using these first cycle coding methods, I began to analyze the list of over one hundred codes for patterns. As a second cycle coding method, I used pattern coding, in order to group the first cycle codes into fewer categories (Saldaña, 2016). From my second cycle pattern coding, I noted five

emerging themes. The emerging themes and subthemes formed the basis of my narrative analysis.

As I engaged in second cycle pattern coding, I sought to identify any patterns that fell under the umbrella of the three sensemaking subprocesses, in non-linear order (Coburn, 2001): taking cues from the social establishment, taking cues from the institutional establishment (including but not limited to understanding grading frameworks and expectations), and interpreting and adapting or resisting grading practices. The latter was the most relevant to the emerging themes. After coding, categorizing, and reflecting on the data (Saldaña, 2016) in these ways, I themed the data. According to Saldaña (2016), a theme is a phrase or sentence that serves to pinpoint the meaning of a unit of data. The ultimate goal of theming the data is to arrive at an “overarching” theme of all of the data or a theme that integrates the other themes together (Saldaña, 2016, p. 199).

Document analysis was a multi-stage process, involving identifying categories and qualities, interpreting social processes, and contextualizing the document (Hodder, 2012). Documents that were considered relevant for this analysis included course syllabi or any document reflective of the teacher’s grading criteria or grading practice in any way. The document analysis was guided by a protocol (see Appendix D) and questions and emerging findings from the interviews. In analyzing the documents, I sought to understand the information that is being communicated, implicitly and explicitly. In the same way I coded the interview transcripts, I use concept coding as a first cycle method for document analysis, followed by pattern coding as a second cycle technique. Concept coding is an appropriate method to use because it implies ideas rather than observable

behaviors or spoken words (Saldaña, 2016). The concepts noted will be nouns, based on what is conveyed in the document and what is left out. As I engaged in document analysis, I was attuned emerging patterns with regard to the subprocess ‘taking cues from the institutional establishment’ by evaluating what information was communicated with respect to how teachers reported their actual implementation of stated policies.

Throughout the data collection and coding processes, I took steps to ensure that the data was organized. As a part of the organizational process, I utilized matrices in MAXQDA to document and store the data based on research questions, codes, or categories (Rossman & Rallis, 2017).

Ethical Considerations

There were three ethical considerations in this study: First was protecting the anonymity of my respondents. The second and third are more personal in nature: my role as an insider in the district I am studying and my role as a parent of a middle school student in the district I am studying. In order to protect the anonymity of my respondents, I assigned a number to each respondent and to their corresponding documents. In seeking participation in this study, I provided informed consent forms to all participants (see appendix A) in advance of and at the time of the interview. Informed consent ensures that all participants understand the nature of the study in which they are participating, and it is a key requirement of the IRB process (Rubin & Rubin, 2012). Informed consent includes making all subjects aware of any potential risks involved in the study and ensuring that participants are not overtly or covertly forced to participate (Rubin & Rubin, 2012). The informed consent letter explained the research study, the time commitment expected, and my confidentiality policy. I also explained verbally, prior to each interview, that the nature of their participation was voluntary and

anonymous. The study was also carefully vetted by an IRB board through Rowan University and through the Board of Education in the district being studied.

As an insider, both as a parent and as a teacher, I will be reflective of my positionality in this study. Being an insider means operating in the dual role of researcher and practitioner, and while this role offers many benefits—including a deep understanding of the context which one is studying—there are also several potential problems in this dual role that must be addressed (Herr & Anderson, 2015). As an insider, I was researching my own practice setting, and in some situations, I interviewed teachers with whom I have existing collegial relationships. Yet, I do not hold a position of power—I am a teacher who will be interviewing other teachers. My role as an insider made me more empathetic to the desire for anonymity, and I assured that each of my respondents remained anonymous. I was also methodical about disclosure; any respondents with whom I am not familiar were fully informed as to my role as an insider, the nature of the study, and their rights as a participant. To guard against my personal professional knowledge or opinion being included to sway or influence the interviewees or the direction of their responses, I was careful to “bracket” and hold off on any information or discussion that will prevent my participants from exploring and explaining their experiences in their own terms (Stringer, 2014, p. 106). I was also careful to separate my roles as parent, teacher, and researcher. Considering the potential for my own preconceived notions as an insider, I piloted my research questions to other teachers, and was careful to avoid speculation or presumption in the research questions. I also asked that the teacher-evaluators look for potential biases.

In addition to ethical concerns pertaining to the participants and my role as an insider, as a qualitative researcher, I must also be aware of my own perspectives and biases (Rossman & Rallis, 2017). Prior to beginning this research project, I reflected on my potential biases. One potential bias I am aware of is derived from my own teaching practice. I know that I prefer the role of advocate over the role of judge, and I fear that I am sometimes lenient in grading struggling students. I entered this research understanding my biases, and I have brought them to the forefront of my awareness as a researcher. In addition to reflecting on my biases, I clarified my worldview, which will inevitably influence my research. The four worldviews share the same basic elements (ontology, epistemology, axiology, methodology, and rhetoric), yet filter experiences through those elements very differently (Creswell, 2009c; Lincoln & Guba, 2000; Creswell & Plano Clark, 2011). This research was conducted under the worldview of constructivism, which derives understanding informed by the subjective views, derived through interviews with multiple respondents (Creswell & Plano Clark, 2011). This type of research presumes that respondents experiences and perceptions are informed by their social context as well as their own personal experiences (Creswell & Plano Clark, 2011). Characteristically, any patterns or understandings derived from this type of research emerge as the research is conducted.

Institutional Review Board

Prior to any research being conducted, a research proposal was be submitted to both the Rowan University Institutional Review Board (IRB) and the Board of Education (BOE) of the district that is the foci of this study. According to the Rowan University website, the mission of the IRB offices has four key components: to protect any and all

participants in research studies, to ensure compliance with federal and state laws and regulations, to protect and ensure the ethical treatment of human subjects in research, and to offer education to the University's researchers about regulations and best practices. The respective offices—the Rowan IRB and the BOE—received a proposal of the study, including sections explaining the purpose of the proposed study, a literature review, and an explanation of the methodology. Both offices reviewed all proposed research studies and then offered approval for the research study to be conducted.

Limitations

Discussions on grading are multi-faceted and often controversial. Aside from best practices, there has emerged a variety of discourse on grading ranging from abolishing grading, to ending the practice of assigning zeros, to using standards based grades. Given the multitudinous array of discussions on the broad topic of grading, this study is limited to a discussion on teachers' perspectives on grading practices with regard to writing, specifically. Given the nature of the questions in this study there will be some questions about grading that this study will not address. This study, for example, will not answer questions about grade inflation or the ungrading movement. This study is also limited to examining the practices in a single district, a specific subject, and a specific content area. Furthermore, this study does not examine other elements of grading within this subject that may be interesting for future research, such as grading in resource room or LLD settings.

Summary

In conclusion, this qualitative case study will examine sixth through twelfth grade language arts teachers' perceptions with regard to their grading practices. Through

qualitative interviewing and analysis, this study will seek to identify specific elements of the grading process that are shared or unique to each context. This study will look at teachers' descriptions of their grading practices in order to identify features of teachers' practices, to identify if and how pedagogy and curriculum influence the grading process, and to understand how teachers' grading practices are affected by grade-level, classroom context, or individual teachers' beliefs.

Chapter 4

Findings

This chapter presents my findings from the analysis of the qualitative interviews and document analysis conducted for this study. This chapter will be topically organized, with data presented thematically, in relation to the research questions this study sought to address:

- What contextual factors inform middle school and high school language arts teachers' grading/ assessing practices for writing?
- What are teachers' values and beliefs about the age group they teach, and learning and motivation in general; and, how do these beliefs influence their writing grading practices (adapted from McMillan, 2019)?
- How do middle school and high school language arts teachers make sense of their writing grading practices?
- What are teachers' experiences with grading writing in their respective classrooms?

To address these questions, I conducted nineteen qualitative interviews with seven middle school teachers and thirteen high school teachers, across five schools. I also collected documents and field notes. This chapter will describe the participants and setting, outline the data collection methods, and present a narrative description of the themes and subthemes that emerged from the data.

Participant Sample and Setting

Each of the nineteen teachers who were interviewed for this study described their grading practices as being informed by either formal training or experience. Six teachers noted formal training that included one or more of the following: apprenticeships, observation, collaborative writing/ grading, creating rubrics for assessment, college or master's level classes, or meaningful professional development. The majority of respondents indicated that while they had no formal training in grading writing, they had carefully honed their grading and assessment practices through experience. The experience they described ranged from diving in and grading essays, working and talking with colleagues about grading, reading professional literature, using rubrics, and one respondent noted that her practice was largely informed by her mom, who helped her cultivate her own skills in a way that is reflected and adapted in her own practice as an educator.

Though this study only examines how teachers make sense of their grading practices with regard to students' writing, it should be noted that the teachers who participated in this study had a range of other instructional duties. They also had a wide range of learners. Middle school teachers had four sections of language arts, and all but one teacher taught classes with differing instructional levels, including: general education, in class support (ICS), and honors. High school teachers (following an A B schedule, meaning they met with different groups on different days) had six course sections, including mostly language arts classes, but in some instances also elective classes, such as journalism, SAT prep, or creative writing. High school teachers also taught different instructional levels, including general education, in class resource (ICR),

and honors. Six teachers taught differing grade levels, in addition to differing instructional levels.

Table 1

Participant Demographics

Middle school		High School	
Teachers interviewed	7	Teachers interviewed	12
Average years teaching	8.7	Average years teaching	10.6
Range of years teaching	4-16	Range of years teaching	4-20
Average number of students	95	Average number of students	133
Student population range	82-107	Student population range*	75-162

(*Some teachers indicated the number of language arts students they teach, excluding electives)

The State of New Jersey Department of Education data from the 2017/ 2018 school year shows that the population of the three middle schools where these interviews were conducted ranged from 650 to over 1,000. The population of the two high schools was approximately 1,360 and approximately 1,420 respectively. It was noted in several interviews that class sizes have increased in recent years.

Data Analysis

Data analysis began with transcribing all interviews and using first cycle coding methods on the transcriptions, the documents, and the field notes. The transcripts were coded using a qualitative analytic software program, MAXQDA Analytics Pro 2020. The initial coding resulted in a total of 744 codes, with approximately 100 unique codes.

Initial codes that appeared across interviews with a frequency of ten or more are presented in the table below.

Table 2

Frequently used codes during initial coding (codes with 10 or more references)

Rubric	44
Setting expectations	27
Perceiving time constraints	22
Taking cues from students: motivation	21
Discussing time	20
Teacher values: expectations	19
Teaching to a range of students	18
Focusing on specific skills	18
Taking cues from colleagues	17
Suggesting potential misalignment	15
Thoughts on marking up the paper	14
Considering other factors: completion/ effort	13
Teacher perception: best practice	13
Noting challenges in assessing reading through writing	12
Discussing what's ahead	12
Offering feedback	12
Taking cues from students: Receiving feedback	12
Teacher values: expressing self-perception	12
Perceiving autonomy	12
Assessing informally [as distinct from formatively]	11
Training by experience	10
Adjusting to change	10

Themes in the Data

As a second cycle coding technique, I noted categorical patterns amongst the initial codes. In order to organize the data further, I analyzed the patterns into themes and

subthemes. The themes will be discussed in five sections in this chapter: time, considering motivation, setting expectations, feedback, and responding to context. Sub themes will be presented in each category, as well.

Table 3

Pattern Coding, Themes

Theme	Codes
Time	perceiving time constraints, discussing time, focusing on specific skills, assessing informally, adjusting to change
Considering motivation	taking cues from students: motivation, considering other factors: completion and effort, discussing what's ahead
Setting expectations	rubric, setting expectations, teacher values: expectations, teacher perception: best practice, taking cues from students: receiving feedback, training by experience
Feedback	suggesting potential misalignment, thoughts on marking up the paper, noting challenges in assessing reading through writing, offering feedback
Responding to Context	teaching to a range of students, perceiving autonomy, taking cues from colleagues

Time

Reading, invariably, takes time. Reading students' writing with a critical lens—for the purpose of grading or offering feedback—is particularly time consuming. With this basic premise, it is not surprising that a predominant theme that emerged consistently across all interviews with regard to grading students writing was time. Several teachers quantified the time they typically allot to grading essays. One high school teacher noted, "...I spend like 20 minutes per essay. I know everyone can't do that" (High School, Teacher 8). Another high school teacher reported, "I mean, I think my five paragraph

essays, I have down to eight minutes an essay, which multiply that by [115+]. It is quite a lot” (High School, Teacher 1). This equates to roughly 16 hours of grading time. One middle school teacher also reported, “I would say I spend at least five to ten minutes on a paper. It takes me a while” (Middle School, Teacher 1). These teachers observations, although varied, all equate grading with time.

Before moving on to a discussion of teachers’ perceptions of how time affects their grading practices, two ideas must be established. First, teachers associated the number of students under their purview (i.e. the volume of writing teachers were responsible for assessing) with time constraints. Second, teachers perceived these time constraints as limiting to their intentions and their grading practices in a variety of ways. In addressing the first idea, one teacher expressed the association of class size and time constraints explicitly and succinctly, as he described the predominant impediment to best practice in grading students’ writing, “Class size, and then that turns into time” (High School, Teacher 4). Another high school teacher exemplified how time constraints can be perceived as limiting teachers intentions/ practices, “There's so many things that I would like to do, but having so many of them [students], it takes me so long to grade them that by the time I give them back, I'm already onto something else” (High School, Teacher 14). As this teacher’s words suggest, time can limit a teacher, or a teacher can limit time. This quotation is an appropriate jumping off point for my discussion on time, as teachers were frequently making decisions between those two things. No teacher claimed to have a perfect system; they all acknowledged some form of limitation in their grading practices, yet these limitations manifested in different ways. Teachers alternately discussed ways in which they perceived they were limited by time constraints and ways

in which they adapted their practices so that they limited the amount of time they spent grading students' writing.

Perceiving limits with timeliness. Teachers' individual grading practices were inevitably shaped by their intentions, their school context, and sometimes by their personal lives. The majority of teachers explicitly noted their intention was to get feedback to students quickly, yet they grappled with time constraints and high grading volume. Several high school teachers juxtaposed their struggle to grade in a timely manner to their intention as teachers. This juxtaposition was apparent in the following response, "So, the number of students greatly impacts, I would say, the way that we grade their writing. Just because you're not able to collect something and hand it back the next day like you would want to" (High School, Teacher 6). Most teachers' discussions about time constraints, such as this one, lamented the barriers to offering students grades or feedback quickly. This teacher's intention may have been to get students feedback quickly, yet this was an impossibility in light of other factors.

Many teachers discussed how grading demands inevitably went beyond the scope of the school day. For some teachers, timeliness suffered because of personal demands that impeded timely grading in personal contexts, as well:

...there's just so much work on every front that the timeliness suffers, and I think having 90 kids and if they're all writing a paragraph, let alone an essay, that's a lot of time to spend on each physically. Sometimes just as I said, as a mom, and as someone commuting, and working full time, sometimes I'm like, I literally don't have another hour in the day... (High School, Teacher 2).

Several other teachers discussed coaching positions, parenting responsibilities, and personal time. Though these teachers consistently expressed a desire to get feedback and grades to students in a timely manner, they often described making sacrifices to their personal schedule to do so. Throughout the interviews, it was abundantly clear that grading writing is demanding not only of teachers' time but of their care and attention.

Perceiving limits: Doing justice. Teachers often implied that grading students work was not only time consuming, it was intellectually consuming, as well. Teachers alternately discussed how grading students writing required focus, care, and attention. Several teachers noted how—in light of time constraints—they sometime made sacrifices in the aforementioned categories. One teacher describes, “I just don't physically have the time to do the amount of grading that I did before.” She elaborated, “Because it's [150+] students, whereas two years ago it was 106 students. It's like adding two more classes to the mix, and that certainly cuts down on the stuff that I can actually grade carefully” (High School, Teacher 7). The distinction between grading to get it done and grading carefully is an important one. Teachers often expressed issues such as this, which suggests what they viewed as best practice was impeded by reality. One middle school teacher reported:

It's definitely difficult. I don't think I give every paper the justice that it deserves. So, I try to spend as much time as possible on each paper, but I still have to keep in consideration that I have to grade the papers within the given amount of time to give them back” (Middle School, Teacher 1).

This particular teacher's reflection reveals a collision of value systems experienced by the majority of the teachers interviewed. Teachers alternately had to decide between making

sacrifices in terms of “justice” to each paper—in terms of giving it the attention warranted for a careful grade and/ or feedback—and in terms of getting the work back within a time window that still bore relevance to instruction and to the student. Teachers often expressed feeling like they wanted to do justice to grading or assessing students’ work, yet they were grappling with finite time. Some teachers opted to sacrifice timelines in favor of careful grading, some teachers opted to sacrifice personal time in favor of timeliness, and other teachers had variations therein. In these decision making processes, there was a constant trade off of values, intentions, and reality. Moreover, these were decision that seemed to be made in flux, as the nature of the variables was constantly changing. Teachers’ intentions, personal lives, and classroom realities created a dynamic interplay of factors, each of which influenced the grading decision they made. As a result, there were variations in what they chose to grade and how they chose to grade it.

Mitigating time constraints. Many teachers discussed limiting what they graded based on scope or specific skills, for specific assignments. One teacher noted, “Well, I have so many students, so when I grade their writing I try to focus on the specific skills that we’re practicing at the moment. Otherwise I’ll get caught up fixing too many things at once” (High School, Teacher 11). Another teacher reported, “Sometimes I’ll grade just excerpts of essays, like the first time my students write an introduction, that’ll get a grade. So small pieces like that” (Middle School, Teacher 4). Finally, a high school teacher described how the volume of grading impacts her practice:

I have [approximately 120 students], I can’t possibly spend the time necessary really to I’ll look for maybe one or two skills only, leading up to a bigger

assignment grade every single thing they do. So mostly it affects the assignments that I choose to grade. Sometimes I'll have them do a series of assignments and then have them reflect and I'll choose, or they'll choose, or I'll choose one and I'll focus on maybe one of the series. Sometimes we'll have a few small writing assignments that (High School, Teacher 1).

Other teachers similarly mentioned maintaining the volume of writing students did, while limiting the scope of what teachers actually graded. Generally, these teachers limited what was graded through student-selection practices, such as the one described above. Student selection was described as a practice whereby students review all of their work in a specific category, and then choose the assignment on which they wanted to receive a grade. Several other teachers indicated utilizing this practice. A middle school teacher discussed this practice, as well. After describing a first round of grading, where she may only grade a specific paragraph selected by the student, she described her intention in personalizing her grading practice to her students' goals. She described allowing her students' goals to dictate what she grades. If, as in the example she selected, they chose to increase their use of transitions as a personal goal, then her grading would focus on that goal, "And then I'll just grade them based on their transition use that fits their personal goal" (Middle School, Teacher 7). While this teachers' response is representative of how teachers limited their grading to specific elements, other teachers took a different approach.

Some teachers noted that they limited what they would actually assign to students to mitigate their grading volume. Limiting what they would assign to students suggests an embedded belief that teachers must grade all work assigned. One teacher expressed

the practice of limiting what was assigned in terms of a tradeoff. In order to maximize the value of the feedback he was able to offer, he had to limit the volume of the work he assigned. He describes, “If I get it back to them very quickly, it may not have anything useful on it except for a number and maybe a few comments. I end up spacing out their writing, so they are writing less than they have been” (High School, Teacher 4). Another high school teacher similarly noted restricting the amount of work assigned:

The amount of kids doesn't really affect it [my grading practice]. It's how much I can actually assign, which I feel like I've been changing some of my assignments because there's just not enough time in the day (High School, Teacher 8).

Some teachers, such as this one, suggested that their high-integrity grading practices remained intact, yet in order to maintain their thorough grading practices, what they assigned to be graded was instead affected. High school teachers' had more students under their purview (almost 50% more, in some instances). Some teachers discussed how their grading practices were affected in terms of shifting responsibility to their students in terms of seeking feedback. This high school teacher expresses how she correlates more students with more personal responsibility put on them. In describing increasing class sizes, this teacher noted:

With less kids in the class, I was able to have more drafts of an essay, meet one on one with every single one of them, and check in on their progress a little bit more. Now that's much more difficult in that it puts the onus on them to come to my extra help time and to see me outside of class if they want me to look over a draft, and some of them don't do that (High School, Teacher 12).

Though the shift in responsibility that this teacher notes may be developmentally appropriate, it is directly informed by contextual factors rather than developmental needs. Aside from nuances such as this one, teachers' responses in terms of feeling impacted by time and grading constraints were consistent across contexts.

Considering Motivation

Several teachers explicitly stated or alluded to their perceptions of what their students would find motivating and what their students would potentially find demotivating. These perceptions were expressed with regard to grades and feedback.

Perceptions of motivating factors. High school teachers expressed perceptions that grades are the nexus of student motivation. They perceived that their students produced better quality work if they knew that their work would be graded. The topic of grades as a motivating factor was present in several high school teachers' responses. In the following quotation, one teacher reports a conversation she had with her colleagues, wherein they discussed the need to give a completion grade for formative assessments to incentivize students completing the work. As she notes, a completion grade was problematic insofar as students did not put in as much effort because they were not being graded in terms of 'correctness.' She described, "...we want to be fair and upfront with the kids and tell them the purpose for our grading, but if we're not counting everything exactly as what they get correct or incorrect, then maybe they're not taking it seriously" (High School, Teacher 10). This conversation suggests that grades may, at times, be utilized as a tool to influence students' behavior. This teacher reports an association between grades and effort, and she is actively engaging with colleagues to try to find a grading system that ensures student effort. This collegial discussion about grading also

suggests that teachers are actively and collectively grappling with grading issues that affect their practices.

One other manifestation of teachers' behavioral perceptions of student motivation was expressed in terms of perceived leniency. One teacher noted his challenges with regard to the categorical weighting of 'major' and 'minor' assignments. He said, "I definitely know they tell me that I am way tougher on their essays than on the smaller writing assignments, so I'm working on that. I'm like, well I guess I could be a little tougher on the smaller assignments" (High School, Teacher 9). With that said, high school teachers, such as this one, often alluded to the idea that students benefited more from a tough grader. The general perception was that students were more likely to rise to the expectations of a tough grader. Again, this has behavioristic implications, as it implicitly posits the grade as a controlling factor of desired academic enablers (effort and motivation). This teacher's discussion of behaviorism was consistent with elements of several other high school teachers' discussions of grading and work completion. Another teacher, for example, observed, "... grading is so subjective that the kids get to know in that first marking period what the expectations of the teacher are, and they either cater up or down to that" (High School, Teacher 7). This teacher's observation suggests behavioristic implications, as well. She suggests that students motivation and effort is dependent upon and gauged to their perceptions of the teacher's expectations. Though some high school teachers also expressed humanistic/ constructivist intentions, these intentions were clearly limited by contextual factors. The middle school teachers collectively presented a more consistently humanistic approach.

Perceptions of safeguarding students. The majority of middle school teachers expressed or alluded to constructivist or humanistic approaches. Teachers who aligned themselves with constructivist pedagogy often expressed wariness of anything that had potential negative implications to students' self-esteem. In some cases, teachers connected their practices to the curriculum ascribed to them. One middle school teacher, for example, explicitly made the association between the workshop approach and her intention of limiting feedback on a final paper. She described how her practice was aligned with the workshop model, "We're taught to have students fix mistakes before they turn something in... In workshop, you're not supposed to leave a lot of comments on their papers. It's really based off that rubric and that checklist" (Middle School, Teacher 5). Another middle school teachers noted, "I think some of the rubrics they have a lot of, it's just a lot. It's hard to cover without crushing students. We also want to build their confidence, so I try to modify it a bit. (Middle School, Teacher 3). Beyond considering assessment tools, other teachers perceived students' feelings in receiving grades or marked-up writing. One teacher offered this advice, "Adjust your expectations and your 'drive to revise.' You can crush a student with a marked-up essay. To them, a grade is final, done... they may not want to revise" (High School, Teacher 5). In unpacking this quotation, there are several underlying beliefs. First, this teacher suggests that students are attaching feelings of self-worth to teacher feedback. This teacher also suggest that she anticipates that students may view the grade as final. In each of these instances, teachers were basing their actions on safeguarding students in some way, and each of these teachers expressed pedagogical beliefs that favored constructivism and humanism.

Setting Expectations Before Grading

In discussing how they set expectations, teachers' responses represented their understanding of both their instructional age group and their views on best practice in grading students' writing.

Setting expectations. Analytic grading means that feedback is offered based on specific criteria in a writing assignment. This form of grading is most readily associated with rubrics, and rubrics were the most common assessment tool used in both grading and in setting expectations with students prior to grading a writing assignment, across interviews. This theme discusses teachers' descriptions of utilizing and perceiving rubrics as tools for setting expectations, as well as tools for assessing students' writing. The majority of teachers responded that they use the rubric in some way to set expectations. Middle school teachers' *Units of Study* include rubrics for each unit and for each summative writing piece. Middle school teachers were aligned in their practice of utilizing and often adapting the rubrics; their collective participation in this practice suggests that these decisions were social ones, rather than isolated anomalies. High school teachers did not discuss any prescribed rubrics; instead, they described eclectic approaches, including creating rubrics from scratch, modifying rubrics found online, and utilizing standardized rubrics for a variety of assignments.

It can be gleaned from teachers' discussions about grading writing that before the writing process ensues, teaching involves showing students the expectations. In that regard, the rubric was discussed as a tool for both reinforcing expectations and for grading. One high school teacher noted:

We all are just responsible for creating our own rubric, so I usually... I don't like to surprise them at all. The students will know exactly what's on that rubric. I usually design it based on whatever they are studying or an accumulation of things we've already studied. So, there are some things that appear on every rubric that aren't a surprise, [like] writing conventions (High School, Teacher 1).

This teacher offers insight into several important factors with regard to setting expectations at the high school level. First, this exemplifies how teachers prioritize a student's understanding of how they will be graded. Second, it shows variation in assessment, from assignment to assignment, and—it can be inferred from the flexibility suggested here—from teacher to teacher and grade to grade. At the high school level, the rubric seemed to be used as a tool for communicating as much as for grading. Several high school teachers discussed creating their own assessment tools, predominantly rubrics. In response to a question about why she choose to go through the extra work of creating her own rubrics, one teacher responded that she did so to obviate the problem of “vague” rubrics she was finding online. She elaborated on the problem vague rubrics caused, “And then when the students asked me, ‘Well, why did I score here instead of here? I wasn't able to answer them. And so, if I can't answer them, then why did I score them there instead of there?’” (High School, Teacher 11). In this way, setting expectations prior to grading was presented as equally important for the teacher as it was for the recipient of the grade. Teachers, it seems, like clarity around expectations for grades, as well. The following middle school teacher's response represents the common practice amongst the middle school teachers: “We always go over the rubric ahead of time, and I usually have them self-assess with that rubric beforehand and reflect... I want

to be transparent, from the start, what I'm looking for specifically with grading” (Middle School, Teacher 2). Though the majority of teachers discussed practices that involved clear expectations in advance of major writing assessments, there was more collective agreement and alignment amongst the middle school teachers, based on their responses.

Expectations and the instructional age group. Setting expectations has practical implications. It can set the bar for student achievement, communicate learning expectations, and obviate confusion. Middle school teachers’ expectations were presented as fluid, calibrated to their learning groups. High school teachers’ expectations appeared less flexible. One commonality that appeared in four of the high school interviews that did not occur at all in the middle school interviews was teachers alluding to the idea that certain expectations were “implied” (High School, Teacher 1). Or, expectations were released to students. Two teachers offered students an overview of the entire course of instruction at the onset, “They automatically know what to expect from here, from my syllabus. I have everything broken down...they know exactly what is expected of them” (High School, Teacher 8). This teacher’s syllabus was submitted for document analysis, and it is comprehensive in scope. It features a calendar of assignments for the school year, through March. The grading policy is clearly articulated with categorical weighting and specific assignments. This document reiterates this teacher’s emphasis on setting clear expectations. Another teacher’s response couches classroom policy within broader departmental policy, as she explained that the high school teachers followed a 60% / 40% split for “major” and “minor” assignments. Another teacher relied on systemic communication. When asked how she communicated her grading policy to her students, she responded, “Well, students know the new grade

breakdown for the grade book” (High School, Teacher 12). This presents the idea that high school teachers may expect student familiarity systemic grading policy, in addition to their expectations about understanding classroom policy.

Unique to the high school context, students are increasingly perceived as independent and responsible for their own schedules and learning responsibilities. Another high school teacher noted, “It's a breakdown of the rubric or is it discussion just of the expectations. I'll ask my kids at times, are these fair? Is what I'm looking for something you think you can do?” (High School, Teacher 4). I responded with the follow-up question *Do they usually think it's fair?* The teacher's response was, “Either I'll tell them like, ‘Well, it's going to be tough, but you're going to do it anyway,’ or we'll work in something that makes it more realistic for them to succeed but still be challenged.” This response “...it's going to be tough, but you're going to do it anyway...” is revealing of the high school teachers' collective embedded philosophy. Though teachers' responses still suggested actively meeting diverse student needs, the beliefs they suggested were in line with motivating students to rise to expectations.

Middle school teachers did not mention needing to set grading expectations in advance of minor assignments and none mentioned anticipating student resistance to formative or summative assessments. One middle school teacher, for example, described an assignment she planned to add, “It's not in the curriculum but it's a little practice for the standardized test and gives me a snippet of their writing ability” (Middle School, Teacher 3). This represents middle school teachers perceived flexibility in what they are adding to the curriculum. My general perception was that high school teachers

perceived—and took measures to obviate—student grievances in the grading process. This was not a theme that emerged in the middle school context.

Offering Feedback

Feedback was offered in a variety of ways, which were not necessarily mutually exclusive: in conferences, on paper (or digitally), or via the rubric. Teachers' values and beliefs about grading inevitably had implications on how they graded and offered feedback on their students' writing.

Estimating the value of feedback. Teachers suggested that feedback was best delivered with relevance to instruction. This, however, was not always possible. Many teachers alluded to the idea that the more prolonged their grading process was, the less productive their feedback was as a teaching tool for instruction and to inform students. Several teachers noted that by the time the grade was reported, they perceived that the feedback and the message the grade communicated had lost relevance to the students. Some teachers tried to work around this. A middle school teacher emphasized the importance of timely feedback to writing instruction emphasized her role in providing carefully timed, relevant feedback, "...kids at this age are like goldfish and they're going to forget. So, if you can be quick with your response and feedback, they're more likely to listen to it and it mean anything to them ...” (Middle School, Teacher 4). This teacher also equated timely feedback with application, suggesting that students were more likely to apply feedback to what they were learning if feedback was offered within a window of relevancy. This response was echoed in another teacher's emphasis on timely feedback: “I think timely feedback and consistent or frequent feedback, frequent writing with frequent feedback that's timely. So, it's not like a month has passed since

they wrote the essay and now you're giving them feedback..." (High School, Teacher 2). Yet, in this instance, the teacher reported struggling to meet her own perception of best practice with regard to timely feedback. In this, she was not alone.

What teachers reported as best practice often conflicted with what they were able to do with high fidelity in the classroom, given contextual constraints, such as time. In explaining how she anticipated students' reactions to feedback on a graded paper, one teacher alluded to both the utility of her time and to the futility of marking up a graded paper. She noted, "...I think spending your time marking up a paper for every single error for every student, it's number one, impossible and number two, a waste of time because... most of them are like, 'Okay, that's my grade C? Cool'" (High School, Teacher 6). This response is important as it reveals intuitive calculations of the value of a teacher's time.

Lean feedback. When it came to feedback, many teachers discussed intentional limits they placed on the feedback they would offer. There were three iterations of lean feedback offered by teachers: feedback that was limited to specific skills, feedback that was limited to content, and feedback that was limited based on a teachers' perception of digestibility or utility for the students. Several teachers mentioned limiting feedback to specific skills, at least on certain assignments. One teacher expressed her position, informed by advice she received from a professor: "...sometimes we might go through and pick one or two skills that we're really going to comment for a particular round of essays" (High School, Teacher 10). Another teacher focused on quality of thought and content:

I try to look at it almost like I try to ignore all the grammar, spelling, punctuation errors and not get caught up in that. I try to just think, think do they have a clear

claim, an arguable thesis statement, and then do they follow it throughout. So definitely the content and quality of the argument over the actual writing” (High School, Teacher 12).

This response expresses a value system that favors quality of thought and content in students’ writing. Though this is undeniably an important aspect of students’ writing, it is not consistently the favored aspect that is being evaluated by other teachers, or possibly even by this teacher for all assignments. Some teachers favored limiting comments, and typically, they noted intentionally leaving a positive comment in addition to a constructive comment or two. One teacher discussed limiting feedback offered to supplement the rubric in the following way, “Not go over everything bit by bit but go over two things for them that they're doing great and two things that they can improve on” (Middle School, Teacher 1). This teacher’s decision is based on her perception of isolating her feedback to an accessible and digestible message to her students. Overall, teachers’ grading practices had a variety of iterations within individual classrooms, for different assignments, vertically, and horizontally.

Analytic feedback: The shape-shifting, ubiquitous rubric. Two methods of grading/ offering feedback were predominantly discussed: holistic and analytic. The idea of holistic versus analytical grading manifests an important distinction in terms of the totality (or lack thereof) of what is being assessed. Holistic grading implies that students are being graded on the overall quality of writing, not just on isolated skills or categories. High school teachers who choose to grade holistically were inevitably grading elements of expression that were not explicitly taught that year. Teachers’ perception of analytic

grading usually emphasized isolating and assessing what was taught. Analytic grading was the predominant grading method discussed across interviews.

Middle school teachers discussed analytic grading with reference to summative writing assignments. Middle school teachers' rubrics were ascribed through the curriculum. Yet, the majority of middle school teachers noted adapting rather than adopting the rubrics ascribed to them. As noted in the section on setting expectations, high school teachers had eclectic approaches to rubrics, including creating them, modifying ones found online, and utilizing standardized rubrics. The way in which high school teachers operated with rubrics had implications for their teaching and grading, as well. Though the high school approach was less unified, in some ways, it manifested as more intentional with respect to individual assignments. One teacher expressed being informed by the *Understanding by Design* approach, associated with Jay McTighe and Grant Wiggins. The premise of this approach is that the end-product expectation informs instruction. This teacher described, "I look at the end product first, what do I expect from them? ... then what should I cover to lead them to that in order to be successful?" (High School, Teacher 12). In this way, the majority of high school teachers who reported utilizing rubrics did so with consideration to the goals expressed on the assessment tool.

Middle school responses showed that teachers calibrated instruction, feedback, and grading tools to their learning group. In several instances, the middle school teachers alluded to adapting the rubric to account for what was taught or to eliminate what was not taught. One teacher reported:

I feel like I'm able to modify the rubric and checklist to fit the needs of my students and also, what I taught throughout the unit. I do think, if there's a certain

area that I don't touch on, it's not really fair for the students to be assessed on that.

I am able to eliminate and modify from the checklist based on their needs and also what I've taught them (Middle School, Teacher 6).

Another middle school teacher noted a similar practice, "...the rubric that our curriculum provides can be a little bit highbrow, in terms of language and its expectation... See what they can do, and then, you can adjust the rubric accordingly" (Middle School, Teacher 1).

As a tool for feedback, the rubric was utilized in different ways. The most important thing to consider in terms of the rubric as a tool for feedback is that it separated the feedback from the actual writing, and it analytically specified categories in which the student was being evaluated.

Some teachers challenged the implicit value placed on rubrics as assessment tools. Teachers who took issue with the excessive use of rubrics posited the idea that rubrics were limited in addressing all of the features of good writing. One teacher noted, "...there's so much gray area as to what goes into a solid piece of writing and sometimes figuring out if I know they did something really great, but I don't see the wording on the rubric..." (High School, Teacher 1). Another high school teacher noted:

...writing is so subjective. Even when you have the rubric with 80 boxes with all sorts of lots of words in each of them, it's still subjective...I could read it and their [later grade teacher] could read the same thing and we could get in different boxes" (High School, Teacher 2).

Though the majority of teachers favored the use of rubrics, they were clearly not perceived as completely reliable.

Holistic feedback: Marking up the paper. Holistic feedback proved to be a divisive topic. Half of the high school teachers mentioned or alluded to utilizing both analytic and holistic grading practices. Two noted that they exclusively favored holistic grading, which refers to a grading practice that looks at the papers overall quality, rather than a pre-specified criterion. Teachers who supported holistic grading often associated this with “marking up the paper,” and they typically noted that this practice was beneficial for students. One teacher noted the benefit that marking up the paper had on students’ learning. He suggested that after the initial shock of a marked up paper, “Then they kind of get on board.” In this way, he seems to suggest that the marked up paper controls for student behavior. When they see the marks, they exert more effort in submitting a correct draft. As an outcome, he notes, “And I do find that by now, because we’re now halfway through the year, many of the common errors are gone because it forces them to proofread their paper” (High School, Teacher 9). This particular teacher (and one other) showed me a stack of papers on which he had clearly invested a significant amount of time offering detailed and thoughtful feedback. One teacher who advocated for holistic grading also questioned the value of parceling out feedback analytically or in reference only to a particular skill/ skills. She noted that this creates a problem of inconsistency and a lack of understanding of ‘good’ writing. She suggested that if teachers are not viewing writing as a holistic process, then neither are students. She alluded to the unintended outcome of inconsistent grading practices: student disorientation. She relayed the hypothetical student questions, “‘Well, why didn’t you mark that off last time? I’ve always done that. I’ve always done that.’ ... Because major things are getting overlooked... No, unfortunately, they should see red marks all over

their paper ...” (High School, Teacher 8). This shows the implicit suggestion that if something is *not* accounted for on the analytic assessment tool, students may perceive that as tacit approval of errors. Another teacher expressed a similar concern, “Because I struggle with thinking that if I don't grade everything sometimes, they're going to make those same mistakes” (High School, Teacher 11). This teacher reiterates the implication that if a teacher does not comment on a student’s errors on a graded paper, then those errors may persist. Moreover, this points to an embedded belief that problems in writing will perpetuate if students are not explicitly taught otherwise. Holistic grading was presented as the most time consuming grading practices. Teachers who utilized holistic feedback represented their view that this was a beneficial learning tool for students. Some teachers favored this practice yet found it time prohibitive:

...sometimes, probably once or twice a marking period I'll grade an essay from start to finish where I'm literally with my purple pen looking at every punctuation error, spelling error. This year, much less of that just because there isn't the time. I'll literally go through and as I'm grading the essay circle the objects on a rubric (High School, Teacher 7).

In discussing holistic versus analytic grading, there was often an explicit clash of values systems. Several teachers who favored analytic grading practices expressed fears that too many marks on a paper were damaging to students. An earlier quotation exemplifies this attitude most clearly, “Adjust your expectations and your ‘drive to revise.’ You can crush a student with a marked-up essay” (High School, Teacher 5). Yet, teachers who favored holistic grading often did so with a sacrifice to their personal time, insisting that it yielded a payout in terms of student learning. This dichotomy in these opinions is clear, and in

both corners, teachers believed that they were best-serving the needs of their learning group.

Noting challenges in offering feedback: Assessing reading through writing. In the middle school context, teachers referred to narrative assignments and creative writing, but otherwise writing was based on ‘writing about reading.’ In high school, the descriptions of writing assignments (aside from journaling, as discussed in some classes) was almost entirely based on what students were reading. Some examples of writing assignments teachers described as using in relation to reading assignments were summaries, quick writes or long writes, journaling (this emerged as a multipurpose medium), readers’ and writers’ notebooks, and essays. Three teachers made specific mention of challenges perceived in assessing students’ writing when they have not read the book that students are reading.

In this way, several teachers expressed their perception that student-choice had supplanted their ability to gauge the quality and accuracy of students’ writing about reading. With regard to being unfamiliar with the students’ texts, one teacher summed up the sentiment:

I think they're reading their own books and we're teaching them a skill, but they're applying it to a book that no one else necessarily in the room is reading and therefore I'm struggling to assess do they really understand that skill that I've supposedly taught them (Middle School, Teacher 3).

Another teacher noted the struggles in assessing student generated theme statements in text she herself hadn’t read, “...but I'm not as familiar with all the books that these kids read. There's just no time in a day to read as many, as much as I want to” (Middle School,

Teacher 7). This particular teacher's response is reflective again of sensitivity and intention, butting up against time constraints (the impossibility of gaining exposure to the plethora of book titles students' read and write about). One high school teacher thoughtfully discussed the reading curriculum in high school, "...I understand that a lot of the texts that we read, especially in my class, are difficult and way out of the wheelhouse of what some of these kids do..." (High School, Teacher 9). This teacher expanded upon this estimation by discussing how he preserved integrity to assessing students' writing. Rather than exclusively using the essays as a comprehension check, he was able to parcel out the specific writing components being assessed: "You don't have to produce a good [thesis] about [the book in the curriculum], but you do have to produce an arguable thesis, and that's worthwhile for me" (High School, Teacher 9). This is not to say that feedback about the reading is not also offered, but he did imply a dangerous potential of double jeopardy in assessing both comprehension and writing quality simultaneously.

Responding to Context

While middle school teachers tended to discuss responsiveness on an individual or small group level, high school teachers noted barriers to individual instruction based on class sizes. High school teachers accommodated for this in several ways. Some still mentioned trying to confer individually or in small groups, though they were challenged by limited time and scheduling constraints. Some offered holistic feedback on student work. Some addressed collective common issues or errors in students' writing. One middle school teacher discussed formative and "authentic assessment" leading up to a graded essay. When asked if she would elucidate on what authentic assessment looked

like in her classroom, she explained, “Conversations with students about what they're doing. I do a lot of reflection as well, so that gives me a little glimpse into how they think they're doing so I can further assist them in those areas” (Middle School, Teacher 5). This teacher’s responses features responsiveness and language that supports the teachers perception of her role in “assisting” students.

High school teachers, noting constraints (including, time, large class sizes, and scheduling conflicts) to smaller conferences, adopted practices that were responsive to student need, yet amenable to the noted constraints. Several teachers noted basing lessons on “common errors” that they observed in students’ writing. This shows that teachers perceived value in feedback and responsiveness and found ways to generate responsiveness, even with contextual limitations. This practice was exemplified in a document offered to me by one of the high school teachers, titled, “2019-2020 Common Essay Errors.” This teacher categorically tracked students’ responsiveness to instruction, with “Round 1” and “Round 2” categories, which alternately established and followed-up on his instructional lessons. The common errors range from grammatical errors, to style issues, to problems with expression. This example highlights how grades may express a final summation, but many teachers viewed the grading process as an opportunity to glean instructional data. Another teacher described a similar practice, “I just start keeping a running record, log of it so that I can do general observations in addition to the additional individualized feedback (High School, Teacher 3). In this way, teachers remained responsive to their students, in light of contextual factors.

Perceiving alignment. The middle school teachers were more likely to perceive alignment in grading practices within middle school—across middle grades both vertically and horizontally—than the high school teachers. The middle school teacher’s response that encapsulates the sentiment of the majority was this one, “I know that all teachers here do use similar rubrics... I know that each teacher has different expectations and might focus more on other areas, but overall, I do think it will be very, very similar in grading (Middle School, Teacher 2). The nuance in this response is significant, and it reiterates what arose in other sections. The variations in teachers grading practices and grading decisions are manifold, even with controlling factors such as aligned assessment tools for major assignments.

High school teachers’ responses noted the likelihood of inconsistency. This is best encapsulated by this response: There’s not enough consistency. Teachers have different philosophies... middle school, you know, is good with a collective philosophy, but there’s different priorities in high school. It’s less clear in high school.... They’ll need to adjust to different expectations next year (High School, Teacher 5).

Both the middle school and the high school teachers’ responses are intuitive in that they acknowledge that different teachers have different expectations and philosophies, yet as the high school teacher’s response suggests, these differences may be more apparent in the high school context. Several teachers recognized that these differing expectations and philosophies are also apparent to students. A high school teacher noted, “Inflation of grades. I think a lot of times students come to me as the [] grade teacher and they're shocked when they get a C on an essay. They've never gotten a C before. They got all As

last year...” (High School, Teacher 12). Differing grading practices, in this sense, can be problematic for students. Teachers’ collective perception about whether grading practices would remain coherent at the high school level was expressed succinctly, “Depends on the teacher” (High School, Teacher 8). And in a separate interview, “It depends on who they get, which is I’m sure the answer you’re going to see from a lot of people” (High School, Teacher 9). Indeed, it was. While high school teachers did perceive horizontal alignment when they specifically mentioned collaborating with grade-level colleagues, they expressed less confidence in their perception of horizontal alignment. Middle school teachers were more likely to perceive both horizontal and vertical alignment within middle school.

Perceiving autonomy. The majority of teachers across contexts perceived autonomy in their grading practices. When asked if she perceived autonomy in her grading practice, one middle school teacher responded:

Yes, I do. I absolutely do. Because if I have a student who I know and I’m watching push themselves, I will not give them a bumped up grade, but I will grade them based on what I know that they’re capable of doing and the effort that they put into it (Middle School, Teacher 7).

A high school teacher noted:

I think there is a lot of autonomy, but I feel like if I did everything how it’s written in our curriculum, I would not have as much autonomy. I do at times; many times, go beyond what’s there because I feel like that’s more beneficial for the students” (High School, Teacher 4).

In perceiving autonomy, it seems that teachers took cues from administrators, colleagues, and policy, yet ultimately, they also made noted adapting their own policies within these broader confines.

Middle school teachers' perceptions of autonomy may have roots in administrative support. One of the documents analyzed reflects an administrative summary of teachers' collective practices, as reported in a faculty department meeting in the 2018/2019 school year. The document supports teachers' perceptions of autonomy in so far as it dictates weighting of categories, yet not necessarily what is being graded or how. Another document submitted for analysis, represents a collective list of teachers' concerns regarding grading. This document presents a collaborative effort to align practices and navigate grading policy with coherence. This document presents a series of questions and responses. In this document, a question was raised, seeking a guide or precedent on how to weight certain assignments, using the example of a notebook check. The administrative response supported teacher autonomy, "There currently isn't a guide. Providing a guide maybe [sic.] beneficial but should not take teacher autonomy away..." This response goes on to note the importance of teachers being able to assess students thinking and the transference of skills through the notebook. Yet, this estimation belies the reality of grading volume, and the volume of additional writing assessments for which teachers discussed being responsible.

Summary

This chapter presents the data that was yielded in nineteen qualitative interviews across five schools. As these interviews show, there are a variety of contextual factors that influence and shape teachers' grading and assessing practices for writing. There are

also many contextual barriers to what teachers perceive as best practice. There was, at times, a disjuncture between what teachers wanted to do and what they perceived as being realistic with regard to realistic constraints such as time. Overall, teachers view grades in a variety of ways, and they make decisions about the grading process in terms of selecting what to grade, how to grade individual assignments, and how to weight those assignments. Teachers also expressed a divergence of opinions with regard to putting feedback on students' graded drafts. Within the high school context, teachers perceived a potential misalignment of their practices with that of other teachers. All teachers acknowledged benefit in various aspects of the grading/ feedback they offered to students, but none of the teachers perceived a perfect system in terms of their own practice or the unity of the whole grading journey that students are ultimately on as they move through their years of schooling.

Chapter 5

Discussion, Conclusions, Recommendations

This qualitative case study explored how middle and high school language arts teachers make sense of their grading and assessment practices for writing. Consistent with the inductive mode of inquiry, I gathered data from participants and then analyzed this data to form categories and themes (Creswell, 2014). These themes became the findings of this study, and in this chapter, they will be put into context with the existing literature and the research questions. Inductive reasoning posits the resulting theory or explanations as the “end point,” and consistent with that mode of inquiry, the results of this study emerged from the data (Creswell, 2014, p. 65). I began this research study with the understanding that the data would essentially lead the way. I was open to any findings that may have emerged. With that understanding in mind, I approached the data with the following research questions:

- What contextual factors inform middle school and high school language arts teachers’ grading/ assessing practices for writing?
- What are teachers’ values and beliefs about the age group they teach, and learning and motivation in general; and, how do these beliefs influence their writing grading practices (adapted from McMillan, 2019)?
- How do middle school and high school language arts teachers make sense of their writing grading practices?
- What are teachers’ experiences with grading writing in their respective classrooms?

This chapter is organized with thematic sections to address each respective research question. These sections will explain contextual factors that affect practice, teachers' values and beliefs, how teachers make sense of their grading practices, and teachers' experiences in their classrooms. In addition, this chapter will include a discussion of implications for future research, recommendations for practice, and conclusions reached. Though the results of this study may have limitations, the recommendations offered herein may provide insight to language arts departments, language arts teachers, school districts, policy makers, and school administrators. The problem statement of this study was concerned with teachers idiosyncratic practices (McMillan, 2001) and the potential for differences horizontally (in a single grade) and vertically (from one grade to another) and between transitional school levels, specifically, middle school and high school. With that said, the results of this study reveal that there is significant variation in teachers' individual grading practices, in teachers' practices across grades, and in teachers practices across school levels.

Contextual Factors

Middle school and high school teachers all expressed or implied a desire to meet the needs of their respective student groups, yet their sensemaking and their grading practices were affected by contextual factors, including size of their instructional group (equated with time), curriculum, and perceived alignment with other teachers. Middle school teachers responses were more collectively aligned. This consistency could be attributed to unifying contextual factors such as their ascribed curriculum or the fact that they were responsible for single grade instructional grouping. This section will address findings with regard to the following contextual factors: time, curriculum, and alignment.

Time. Across contexts, time constraints arose as an important factor in shaping teachers' sensemaking about their grading practices. Several teachers noted that class sizes have increased in recent years, and these teachers described how their grading practices were affected in light of increased class sizes. Teachers related the size of their instructional group to the volume of grading for which they were responsible. Teachers had varied ways of making sense of the time constraints that emerged from high grading volume; some teachers chose to sacrifice time on their weekends to maintain grading practices. Other teachers made a distinction between grading and 'grading carefully,' while other teachers limited the scope of the work that could be graded, in a variety of ways. Existing literature on teachers' grading practices addresses the fact that tension exists between teachers' internal beliefs and values and classroom realities (McMillan, 2003), such as class size. While the reality of limiting factors is acknowledged, there is not insight offered in existing literature as to how those classroom realities limit practice and influence teachers' sensemaking. This study begins to fill that gap, as it offers specific insight into how the size of a teachers' instructional group informs his or her grading practice with regard to writing assessment.

McMillan (2003) points to the fact that tension exists between teachers' internal beliefs, values and classroom realities, external factors. Consistent with that observation, this study shows a tension between teachers' espoused beliefs and their actual practices, with regard to grading students' writing. The findings presented in this study offer insight into a reality of grading that does not receive exclusive attention in the literature: Time. This study also adds something else to the literature—teachers' expressed that their personal lives also affected their grading practices. Many teachers justified what

they were or were not able to accomplish based not only on constraints within the school day but also with regard to constraints in their personal lives. Teachers' personal lives may be a new category worthy of examination, as it is not a perfect fit for "external factors" that McMillan (2003) establishes to account for "state accountability testing, district policies, and parents" nor is it accounted for under the umbrella of "teacher knowledge, beliefs, expectations, and values" or "classroom realities" (McMillan, p. 36). The omission of considering teachers' lives beyond the school day may be a significant one, as teachers consistently referred to what was or wasn't on their metaphorical plate when they discussed their grading practices, and these discussions were not limited to the confines of the classroom. In this way, grading systems come up against personal realities and classroom realities alike. This finding may only bear relevance to grading students' writing, as it is particularly time-consuming, or it may only bear relevance to instances of high grading volume. Either way, it something that should be considered in examining the unique ways in which teachers made sense of their practices.

Curriculum. The workshop model curriculum of the 6-8th grade teachers, based on the *Units of Study*, has an embedded constructivist and humanistic approach to teaching writing. Consistent with this approach, middle school teachers' discussions tended to emphasize elements of choice, consideration of students' preferences, and consideration of students' intrinsic well-being, favoring approaches that avoided negative implications. The majority of the middle school teachers implied that they embraced constructivist pedagogy, which could be the result of the curriculum assigned to them. The high school teachers were less aligned pedagogically; some suggested constructivist inclinations, others suggested behavioristic inclinations, while others expressed beliefs or

practices that were split between philosophies. The curriculum environment in the high school was harder to pin down, as well. It differed based on course and year. It can be gleaned from teacher comments that the curriculum was written by teacher committees, and one teacher noted that the newer curriculums were “hybrid writer/ readers workshop model” adopted from Kelly Gallagher’s *180 Days* (High School, Teacher 9).

This study may highlight the point that teachers contexts and resources may influence their enactment of learning theories. Constructivist and humanistic approaches, according to Beane, Lipka, and Ludewig (1980) offer flexibility and student input in decision making. Constructivist pedagogy embeds three beliefs: learning is active; learning can be personalized to where students are with specific skills or concepts, thus assessment should inform instruction; and finally, students are involved in the practices of a particular discipline (Jonassen, 1999, as cited in Lindner, 2017). Several teachers who discussed delayed feedback, implied a conflict with their ascribed or chosen teaching pedagogy, wherein there was a discrepancy between the pace of instruction and teachers’ ability to maintain consistent, timely feedback that was relevant to instruction. Teachers who did not get feedback to students as quickly as they wanted expressed how this conflicted with their intention. Linder (2017) notes that as a teaching model, it relies on immediacy and responsive teaching. This was more significant in the high school, where teachers had larger instructional groups. What also arose in the high school context was a challenge to formative assessment: Student effort was perceived as lower when students were aware that assignments would be ungraded. Brookhart, Stiggins, McTighe, and William (2019) note that formative assessment is essential, yet implementing it remains a weakness in many districts. This study contributes insight into

possible challenges to formative assessment. Barriers to formative assessment that arose in this study, at the high school level, were volume of grading and perceived lack of student motivation in completing ungraded assignments.

Another important factor is how teachers used their reflections after grading or assessing. High school teachers reported teaching based on common errors, using data from grading. Middle school teachers were more likely to calibrate their instruction to the students.’ This is also pedagogically consistent with the workshop model, as the heart of the workshop model is assessment in order to calibrate instruction to students’ needs.

Perceiving Alignment. Teachers perceptions of alignment where a product of their social interactions. Wherein teachers were more likely to interact and collaborate with teachers horizontally, and these interactions were correlated to teachers’ perceptions of alignment. If teachers did not express social interaction with other teachers horizontally or vertically, they were more likely to express uncertainty or perceptions of misalignment. This is consistent with Coburn’s (2001) argument that sensemaking is a social process, rooted in teaching contexts, where professional interactions led to collective interpretation and understanding. For the purpose of this study, sensemaking accounted for teachers taking cues from their social and institutional establishment (including establishing and perceiving—or not—grading frameworks and expectations.

With that said, teachers fell into three categories in perceiving alignment: they perceived alignment, they perceived misalignment, or they expressed uncertainty. The majority of middle school teachers perceived alignment, respective of other middle school teachers, but they were less certain of consistency beyond their immediate context. Their collective perceptions implies social sensemaking (Coburn, 2001). The majority of

high school teachers perceived misalignment or expressed uncertainty about both vertical and horizontal grading alignment. Regardless of their perceptions of alignment or misalignment, teachers believed that *they* were preparing students for what was ahead. In this way, teachers' perceptions of what teachers in other grades were doing did not appear to influence their respective grading practices. Teachers were influenced, however, by other teachers' grading practices horizontally. In middle school and in high school, teachers who reported working closely with other teachers at their grade level reported consistent philosophies, sharing of resources, and similar grading practices. This is consistent with the theory of sensemaking, in so far as teachers' were social participants, reflective of their local contexts (Weick et al., 2005). When teachers reported sharing practices or philosophies, it was consistently only with teachers whom they directly engaged. In viewing the social nature of sensemaking (Coburn, 2001), it is not surprising that direct engagement with other teachers inevitably influenced teachers' practices, while in the absence of that engagement, teachers' practices were unaffected.

Issues of alignment, both vertical and horizontal, are important, since grading affects a student's trajectory in a variety of ways. This study contributes the unique understanding that teachers' grading practices are more likely to be influenced by teachers with whom they directly engaged, and it seems that there are more opportunities for that alignment to happen within, rather than across grades. This finding has implications in considering the transition from high school to middle school. The literature suggest that academic problems may be exacerbated during transitional years, such as the transition from middle school to high school (Queen, 2002). This study acknowledges that students' grades are likely to fluctuate during transition years; yet the

implications may be as relevant to teachers' grading practices as to real changes in student performance. The latter may not always be the fluctuating value. This study suggests that when examining a drop in students' grades from year to year, or during transitional years, contextual variables beyond a student's performance must also be considered.

Discussion of findings. This study offers several unique findings with regard to assessing students writing and possible to the understanding of teachers' grading practices more broadly. Teachers at the secondary level who are responsible for assessing student writing mitigate time constraints, sometimes at the expense of their best intentions with regard to feedback and instruction. In addition, teachers' grading practices and their perceptions of time may be too narrowly defined in existing literature. This study offers the unique finding that teachers' perceptions of time, more specifically, time spent grading, are not limited to the classroom. Any study that considers teachers' perceptions of grading must also consider teachers' practices in light of their personal lives, as grading was described to extend beyond the confines of the school day. With regard to the curriculum, when pedagogically inclined, like with the *Units of Study*, curriculum may affect a teachers' pedagogical beliefs. Curriculum may inform and unify teachers' practices in some ways, but it must be acknowledged that some environments may be prohibitive of constructivist pedagogy. That is to say, contextual factors (such as class size) may be hostile to constructivist learning philosophies. Constructivist pedagogy may be misaligned with classroom realities. Finally, when considering alignment context must also be considered. Teachers grading practices are likely influenced by the teachers with whom they have the most contact. This final point relates

to Colburn's (2001) subprocesses of collective sensemaking, wherein teachers construct their understanding through social interaction. Coburn's (2001) observation that teachers mitigated technical and practical details collectively is also relevant here.

Teachers Values and Beliefs

This section addresses the research question exploring how teachers' grading practices were potentially affected by their values and beliefs about the age group they teach and learning and motivation in general. Teachers were influenced by their knowledge—shaped and informed by their education, professional development, training, collegial interactions, and experience. Teachers were also influenced by state and district level external factors. The influence of these factors is dynamic; it would be difficult to parcel out any single factor that collectively outweighed the others. This section will address areas of collective tension that arose as a result of teachers' conflicting values and beliefs.

Inconsistent use of assessment tools. One area in which teachers values and beliefs created tension in their individual and/ or collective practices was in terms of how they adopted or adapted assessment tools. Their use of assessment tools was invariably influenced by their beliefs about student learning and motivation. Middle school teachers in this study all utilized the rubrics ascribed by the *Units of Study*. Most middle school teachers noted adapting those rubrics in responsive, yet varied, ways. There were consistent inconsistencies with teachers creation and adaptation of rubrics; the wide variations may adulterate the intention of the rubric as an objective, criterion based tool for assessing academic achievement. This observation is consistent with the literature. The literature presents a consistent picture of inconsistency with regard to grading. Even

in situations where schools adopt grading policies and guidelines, there is variation in teachers' grading practices (Brookhart, 1994; Guskey, 2009; McMillan, 2001).

Moreover, even in when teachers have a strict scoring criterion, as exemplified in Llosa's (2008) study of the English Language Development Classroom Assessment, teachers do not apply the criteria consistently. This study supports the findings that teachers adopt and adapt policy and assessment criteria simultaneously.

Teachers were consistent with best practice in terms of setting expectations before an assignment. Teachers in this study reported aligning their assessments to instructional goals. High school teachers established expectations for major assignments implicitly or based on rubrics, while middle school teachers established expectations more major writing assignments through *Units of Study* rubrics. In many instances, middle school teachers adapted the ascribed rubrics by eliminating categories that did not account for what was taught. The use of rubrics is in-line with best practice, as what is emphasized is "criteria, consistency, and categories" (Guskey & Brookhart, 2019, p. 26). To this end, it is recommended that teachers use checklists, rubrics, and point schemes, and that they offer model responses and bracketed feedback (i.e. feedback should be bracketed if it does not apply to the categories being assessed). The rubric was used in setting expectations, offering feedback, and in defining what would be graded. Best practice in assessment involves establishing clear goals at the onset of an assessment and articulating learning expectations to the students (Guskey & Brookhart, 2019). In addition, assessments should match instructional goals (Brookhart, 2009; McMillan, 2008, Popham, 2008, as cited in Yesbeck 2011). Teachers in both contexts utilized best practices in setting expectations and aligning their rubrics with instructional goals.

The suggestions of best practice discussed in the literature are expressed with regard to goal setting, feedback, and accuracy (Guskey & Brookhart, 2019). Teachers consistently referred to setting expectations (usually through introducing the rubric), they referred to individual conferences as the gold-standard, offering students individual feedback (though teachers consistently noted barriers to having one on one conferences with frequency). An element of best practice that was not mentioned was accuracy in terms of offering grades that were based exclusively on academic achievement. Though accuracy was an element of best practice that was missing from teachers' responses, this is consistent with the literature. Stiggins, Frisbie, and Griswold (1989) suggest that one reason for grading discrepancies may be attributable to the fact that teachers may not know measurement theory. Though teachers did not allude directly to accuracy in their responses, they did refer to objectivity and consistency. It is plausible that this is a gap in professional learning with regard to accuracy in terms of grading academic achievement only.

Several teachers mentioned specific instances where students put forth tremendous effort yet still performed below expectations. In these moments, some teachers reported intuiting that these students would be negatively impacted by a low grade. These teachers were associating poor grades with the potential for students' effort or self-esteem to suffer. While a good grade is perceived as both a motivator and a reward for effort, several teachers' responses implied that a bad grade carried the potential as a demotivator. Consistent with existing studies, several teachers in this study mentioned using effort as a factor in grading students' writing (Cross & Frary, 1999; Kunnath, 2016; McMillan, 2001; Wiley, 2011). In their analysis of existing literature on

grading, Bonner and Chen (2019) note that, “Currently, there seems to be an unspoken and unaddressed acceptance of varied practices, rather than a shared discourse about why a practice is used, how far it is for all students, and how it relates to learning and motivation” (p. 106). This suggestion has implications for teachers’ practices and for equity in grading. Bonner and Chen’s research showed that “academic enabling in grading had a significant positive correlation with favorable views on constructivism” (p. 69). This association of constructivism and academic-enabling in grading is consistent with the findings of this study, as teachers with implicit constructivist pedagogies or constructivist curricular resources (*Units of Study*), tended to allude to or explicitly mention academic-enabling in grading.

One teacher expressed another perspective on this belief system. This teacher described a general education student with writing deficits. She insisted that he was below grade level, and though his effort was preserved, he had lost out on essential remediation of his deficits. This study cannot speak to the validity of her perceptions that he was a victim of effort-based grading, but it is supported by the literature that effort based grading persists (Kunnath, 2016; McMillan & Nash, 2000; Sun & Cheng, 2013), although it is contrary to best-practice and in some instances harmful to learners. Several studies reported that teachers inflated grades of high-effort low-ability students (Cross & Frary, 1999; Kunnath, 2016; McMillan, 2001; Wiley, 2011). Several teachers did in fact suggest that they considered effort in grading students writing.

There was a marked difference in the ways in which middle and high school teachers alluded to their perceptions of student motivation. High school teachers discussed attempts to increase student motivation through behavioristic approaches to

grading. A review of studies relevant to student motivation, grades, and environment showed consistent decline in students' attitudes toward school in the latter years of high school; it is noted that this decline is perceived as students move from middle school to high school (Eccles (Parsons) et al., 1984). The findings of this study are consistent with that observation; teachers in this study perceived motivation in high school and sought ways to motivate students to perform. High school teachers also emphasized setting expectations and ensuring that students understood the grades they received. The literature supports the idea that at the high school level, there is an increased emphasis on personal responsibility (Queen, 2002). According to Messick's (2002) respondents—teachers, parents, and students—the onus of feedback communicated shifted from communicating with parents to communicating with students directly as a purpose of grading. Consistent with the research, this study showed that teachers emphasized personal responsibility and communicating both expectations and feedback directly to students.

Middle school teachers, on the other hand, consistently referred to meeting the needs of the learning group, responding to students' needs, and meeting the students where the students were. That is, the middle school teachers responses emphasized teachers adjusting their expectations to their learners more so than having learners adjust to expectations. The theme of fairness and perception of what would be beneficial to students, a theme that emerged in Sun and Cheng's (2013) study, echoes throughout teachers' discussions of meeting the needs of their respective learning groups, albeit with varied iterations in the middle and high school contexts.

Discussion of findings. The findings of this study are consistent with the existing literature in three ways. Consistent with the literature, the teachers in this study revealed ways in which their values and beliefs influence their grading practices. In addition, many teachers in this study expressed or alluded to academic enabling in their grading practices. Finally, even when criterion-based grading was utilized, variation in grading practices persist.

Where this study contributes to the literature is with regard to use of assessment tools in writing, as this study depicts an incoherent picture of collective assessment practices. This study extends the conversation on writing assessment and best practice in its implications of excessive, yet varied, use of analytic grading of writing using rubrics. Ultimately, teachers metaphorically and literally close their doors and make sense of their own teaching and grading practices, largely in ways that they feel are most accommodating to their students' academic needs, learning preferences, and perceived ability level. Though teachers consistently expressed the best interest of their students, grading practices varied in terms of approach (holistic vs. analytic), tools used (a variety of rubrics or a variety of adaptations of rubrics), and in terms of what was graded.

This study also extends the conversation established by Bonner and Chen (2019), that learning theories may have implications on grading expectations. Bonner and Chen (2019) analyzed teachers' hodgepodge grading practices through the lens of three learning theories: knowledge acquisition, constructivism, and behaviorism. Responsiveness is embedded in constructivist pedagogy. Responsiveness to the learning group seems to be interpreted by some teachers as moving the grading barometer closer to the learning groups' perceived ability, in addition to meeting students where they are

instructionally. This consideration of student ability may be perceived as equitable practice, yet it also may be questioned in terms of future outcomes. Behaviorism was seen in the way high school teachers attempted to mitigate their perception of student motivation and grades. Both of these learning theories were connected to context. Middle school teachers' learning materials, and ascribed teaching styles, were constructivist in nature. The high school teachers behavioristic adaptations seemed to be adapted to mitigate contextual factors.

Making Sense: Grading Within a Social Context

Perhaps one of the greatest challenges to teachers' objectivity is that teachers are inevitably grading within a social context. They are alternately advocating for and evaluating their students (Ebel & Frisbie, 1991), and there is tension in that role. Therefore, in considering how teachers make sense of their writing grading practices, it is essential to examine how teachers make sense of social cues from their students, colleagues, and their teaching contexts. An earlier section addressed an important finding with regard to collegial cues and perceptions of alignment, so this section will focus primarily on how teachers made sense of their practice with regard to taking cues from the students.

Time as an investment: Measuring the returns on feedback. In addressing this section, I am going to start with a metaphor. Teachers are investors. They are investors seeking the greatest return for their investment. To clarify this metaphor, what teachers are investing is their time, the catalyst for this investment is the student, and their promised profits are learning gains. In this regard, it seemed that teachers were both intuitive and data-driven investors. That is to say, they perceived that low-achieving

students could make significant gains if the right conditions were met. They also perceived that they would get a return on meaningful feedback. They gauged the timing and context of their investments, and they sought to maximize their investment in terms of learning gains. What was not addressed was how this also led teachers to make intuitive calculations of when they would *not* get a return on their investment. Teachers expressed a reflexive perception: they perceived and anticipated their students' motivation to utilize feedback. If they perceived low motivation to utilize feedback, they may offer less feedback. Teachers also perceived a window in which their feedback would be most impactful to student learning, and yet returning feedback within this window was made impossible by grading volume and large instructional groups.

Discussion of findings. There is a gap in the literature in addressing teachers' sensemaking with regard to feedback. Existing studies do not examine the calculations that teachers make in terms of offering feedback. This raises potential issues of inequity in grading, as it raises the idea that if a teacher perceives a low-interest, low-effort, low-performing student will not be a worthy investment that will yield a return on their time, they may make decisions that withhold or minimize feedback to these students. Overall, teachers navigated their grading practices in light of a variety of factors. Teachers made modifications to their grading practices based on several contextual factors. Many teachers suggested that they desired to be consistent in grading their individual students, yet the potential for intuitive calculations in offering grading feedback has important implications.

Teachers' Experiences in Their Classrooms

The final research question explored teachers' experiences with grading writing in their respective classrooms. Teachers' grading practices in the middle school and high school level were intentional with regard to students' developmental needs. Consistent with Queen (2002), there was a greater emphasis on personal responsibility at the high school level, in the form of advance notice and personal management of assignments, greater expectation that students internalize expectations, and responsiveness to students' feedback. Viewed collectively, teachers' sensemaking of their unique classroom experiences offers several notable topics: challenges in assessing reading through writing, meeting the needs of a range of learners, and varying perceptions of autonomy.

Assessing reading through writing. Several teachers raised an important point with regard to the challenges posed in assessing students' writing when it is based on students' reading. One high school teacher noted the challenges of assessing students writing based on reading of a complex class text. Middle school teachers discussed challenges in assessing reading through writing that was based on students' independent reading selections. Though independent reading is supported by the workshop model, teachers reported challenges in assessing students' comprehension of texts that they themselves have not read. Similar to teachers who fear they are offering tacit approval of grammar errors that they do not correct, these teachers feared they may be unwittingly offering tacit approval of comprehension gaps expressed in students' writing about their reading. In addition, assessing students' reading through writing raises questions on grading validity when teachers are grading writing about something they have not themselves read.

Teaching to a range of students. There are inherent logistical differences in teaching and grading different instructional levels. Middle school teachers, though primarily responsible for one grade level, taught different instructional levels. High school teachers similarly reported different instructional levels, including college placement, in class resource, honors, and general education, in addition to teaching electives. All New Jersey teachers are beholden to teaching state standards, and all students are responsible for standardized tests, and the same common grade level assessments. Yet, perceptions of students' ability affected teachers' grading practices in terms of how teachers expected students to reach standards or classroom expectations. Teachers largely perceived autonomy in their grading practices, within parameters. Teachers assumed tacit approval in creating assessment material or adapting curriculum rubrics that existed. Some teachers felt autonomy with the distinction that they were beholden to something (standards, a rubric, or common assessments), some teachers perceived autonomy in so far as no one had told them otherwise.

Discussion of findings. In terms of teachers' experiences with their learning groups and their inconsistent creation, adoption, and application of assessment tools, there is a significant amount of variance in what grading actually looks like in individual teacher's classrooms and across grades. Several teachers noted an adjustment period when students started their classes and several teachers alluded to the idea that students would likely need to adjust once again when they moved beyond said teacher's classroom. Yet, from the learner's perspective, this poses problems with interpretability of grades as true measures of academic performance. In an interview with two high school teachers, who were philosophically most aligned with the workshop model of

instruction, it was still apparent that although they were utilizing the practices of workshop (minilessons, small groups, conferring), contextual demands interfered with their ability to be as responsive as they wanted to be. This is significant because it shows that contextual factors may cede pedagogical intentions.

Limitations

Grading processes that teachers engage in are multi-faceted, as are topics of discussion pertaining to grading in education. Grading is a broad topic, and for that reason, it is important to note that this study is limited to exploring teachers' perspectives on their grading practices. It is also limited by sample size; the sample size of this study may impact the ability to generalize beyond this subject and school district. It may also be considered that this study may have limitations relative to time of year. With grading, it is possible that a teacher's perspective may change depending on the volume of grading he or she has at that moment. If this study was conducted, for example, in the summer, teachers may be more reflective with distance from their grading practice. This study was also limited to a specific topic: exploring teachers' sensemaking with regard to their grading practices for writing, in a single subject, in a single district. It did not address questions about grade inflation or abolishing grades.

Recommendations

This study has implications for future research, the school district, and for teachers and educational leaders. The importance of grades, and issues pertaining to their accuracy and interpretability, cannot be understated. Though this is a single case study, and thus may have limitations, the findings in this study have valid implications for research and practice.

Recommendations for future research. Since this is a single case study, the first possibility for future research is that this study could be expanded and replicated to more settings to make it generalizable. This study may also be replicated with relevance to special education, by exploring grading with specific relevance to in class support, in class resource, and resource settings. One other possibility is that this study could be expanded with quantitative data; a mixed methods approach may yield additional findings.

One of the findings of this study was that the majority of teachers are setting expectations through presenting students with their assessment tools. Some teachers raised questions about these scaffolds being potentially detrimental to students, as students were reportedly reliant upon receiving a checklist and writing only with specific, delineated expectations in mind. This leaves room for a study exploring students' writing productivity with and without these scaffolds. While setting expectations is reported as best practices, teachers in the student questioned the implications of this with regard to written expression.

In addition, this study raises questions about mediums for effective feedback. Best practice in grading establishes the need for objective, criterion-based assessment tools. Best practice also establishes the need for effective feedback in setting learning goals. Yet, with regard to feedback for writing, there is a gap in the literature with regard to scientific studies examining learning outcomes and various forms of feedback with writing. There were several concerns raised about rubrics failing to address the “gray area” of good writing. A study exploring learning outcomes with regard to the various

types of feedback mentioned (holistic, analytic, verbal) may address the merits of each respective form of feedback in terms of students' learning gains.

Recommendations for district policy. The implications of this study may inform policy makers with regard to teacher preparation. This study offers insights into possible gaps in professional learning in three aspects: knowledge of measurement theory, real-world training in grading practices to be considered by teacher prep programs, and transparency with regard to writing as a category.

To address these problems, the school district may offer professional learning that is meaningful and relevant to grading, assessing, and student learning/ motivation. In addition, the district may embed time for teachers to norm their grading practices, assessment tools, and identify specific assessments to be reported in the grade book for each learning unit.

Grading practices must consider equity. To that end, there could be a discussion with regard to how teachers modify instructional tools for different instructional levels, in class support, general education, and honors, for middle school and in class resource, general education, and honors or advanced placement for high school. These assessment tools may be collaboratively created or adapted across grades and schools, with consideration of horizontal and vertical alignment, as well as alignment with New Jersey State Learning Standards. Creating assessment tools should not be left to the discretion of individual classroom teachers, aside from modifications based on students individualized educational plans.

Finally, writing should be reported as a distinct category on high school students' report cards. The need for writing achievement to be communicated exclusively is

especially important for college-bound students who may face the prospect of remedial coursework in college. Guskey and Brookhart (2019) discuss accuracy and composite grades, with the suggestion that rather than reporting a single report card grade, students would be better served by dividing grades into categories. Though their suggestions are more concerned with separating grades in terms of product grades, process grades, and progress grades, their observation about the need for clear categories bears relevance to this discussion.

Recommendations for educational leaders. The results of this study may help educational leaders implement a variety of supports, including smaller class sizes, common grading periods, transitional supports, and/ or community partnerships. In discussing the importance of grades beyond communicating with students or informing classroom instruction, Bonner and Chen (2019) note, “Grades are also significant at the administrative and school leadership levels; they influence school and course promotion, they affect graduation rates and school performance indices” (p. 59). With this in mind, this section will posit recommendations for educational leaders.

The first recommendation for practice that emerges from these findings is class size matters with regard to grading and assessment. Educational leaders should carefully consider placing realistic, feasible limits on teachers’ instructional group size. In addition, they should consider streamlining teachers’ instructional responsibilities at the high school level to a single grade level or to a single instructional level, if possible. As Bonner and Chen (2019) note, teachers’ idiosyncratic practices must be addressed in ongoing discourse. To that end, there must also be coherent and frequent collaboration and articulation, horizontally and vertically, within and between schools in a district.

Educational leaders may establish time for staff collaboration, so that they may discuss issues pertaining to grading and assessing.

Other suggestions involve outsourcing some of the grading, while strengthening community partnerships. Perhaps the district can look into a partnership with a local college or university, where preservice teachers are assigned to specific teachers, where they would both student teach and work as apprentice graders. This would promote objectivity in grading. In addition, this would also lighten the grading burden of individual teachers. Conversely, the college or university would benefit in offering more realistic training in terms of managing classroom expectations, which was notably a problem mentioned by teachers.

Implications for my teaching practice. As a classroom teacher, this research is important to my own practice. Classroom teachers often carry the burden of responsibility for student learning and short-comings within our practice, even when some of the issues we are trying to ameliorate are systemic in nature. In terms of improving my own teaching practice, this research reminded me of the implications of my grading practices in conjunction with the students' academic journey. If as teachers, we think of ourselves as siloed in practice, or we place too much emphasis on the importance of our own work in our own individual classrooms, we lose the power of collective impact on students. This study was a powerful reminder to me, as an individual classroom teacher, that I must pursue a collaborative approach to my students' academic journey. Teachers must be individually reflective, yet there must also be opportunities for reflection on our practices horizontally, and vertically, as well.

Conclusion

This single-case study explored teachers' sensemaking with regard to their writing grading practices. Using an inductive mode of inquiry, this study generated findings that addressed gaps in the literature and raised questions for future research. In several areas, the findings of this study were consistent with existing research on teachers' grading practices, but with its exclusive focus on teachers' grading practices with regard to writing, this study offers insights that have not previously been addressed in the literature. This research study offered insight into how teachers make sense of their grading and assessing practices for writing.

References

- About AP Scores. (2016). Retrieved October 12, 2016, from <https://apscore.collegeboard.org/scores/about-ap-scores>
- ACT study points to gaps between Common Core Standards and college expectations. (2016). Retrieved from <http://www.act.org/content/act/en/newsroom/act-study-points-to-gaps-common-core-standards-college-expectations.html>
- Adams, M. J., Fillmore, L. W., Goldenberg, C., Oakhill, J., Paige, D. D., Rasinski, T., & Shanahan, T. (2020, January). Comparing reading research to program design: An examination of Teachers College Units of Study. *Student Achievement Partners*, 1-64. Retrieved from https://docs.google.com/viewerng/viewer?url=https://achievethecore.org/content/upload/Comparing%20Reading%20Research%20to%20Program%20Design_An%20Examination%20of%20Teachers%20College%20Units%20of%20Study%20FINAL.pdf
- Allen, J. D. (2005). Grades as valid measures of academic achievement of classroom learning. *The Clearing House*, 78(5), 218-223. Retrieved from <http://ezproxy.rowan.edu/login?url=https://search.proquest.com/docview/196865980?accountid=13605>
- Altheide, D. L. (2011, December 23). Process of Qualitative Document Analysis. *SAGE Publications, Inc.*, 22-44. <http://dx.doi.org/http://dx.doi.org/10.4135/9781412985536>
- American Federation of Teachers National Council on Measurement in Education National Education Association. (1990, December). Standards for teacher competence in educational assessment of students. *Educational measurement, issues and practice*, 9(4), 30-32. <https://doi.org/10.1111/j.1745-3992.1990.tb00391.x>
- Ampartzaki, M., Kypriotaki, M., Voreadou, C., Dardioti, A., & Stathi, I. (2013). Communities of practice and participatory action research: The formation of synergy for the development of museum programmes for early childhood. *Educational Action Research*, 21(1), 4-27. <http://dx.doi.org/10.1080/09650792.2013.761920>
- Argyris, C., & Schön, D. A. (1974). *Theory in practice: Increasing professional effectiveness*. San Francisco, CA: Jossey-Bass.

- Bailey, M. T. (2012). *The relationship between secondary school teacher perceptions of grading practices and secondary school teacher perceptions of student motivation* (Doctoral dissertation). Retrieved from <http://ezproxy.rowan.edu/login?url=https://search.proquest.com/docview/1011481355?accountid=13605>
- Barnes, S. (1985). A study of classroom pupil evaluation: The missing link in teacher education. *Journal of Teacher Education*, 36(4), 46-49. <https://doi.org/10.1177/002248718503600412>
- Barshay, J. (2019). Scientific evidence on how to teach writing is slim. Retrieved from <https://hechingerreport.org/scientific-evidence-on-how-to-teach-writing-is-slim/>
- Beane, J. A., Lipka, R. P., & Ludewig, J. W. (1980). Synthesis of research on self-concept. *Educational Leadership*, 38(1), 84-89. Retrieved from <http://search.ebscohost.com.ezproxy.rowan.edu/login.aspx?direct=true&db=aph&AN=9394773&site=ehost-live>
- Bonner, S. M., & Chen, P. P. (2019). The composition of grades: Cognitive and noncognitive factors. In T. R. Guskey & S. M. Brookhart (Eds.), *What we know about grading: What works, what doesn't and what's next* (pp. 57-83). Alexandria, VA: ASD.
- Bowers, A. J. (2019). Report card grades and educational outcomes. In T. R. Guskey & S. M. Brookhart (Eds.), *What we know about grading: What works, what doesn't, and what's next* (pp. 32-56). Alexandria, VA: ASCD.
- Brookhart, S. M. (1991, March). Grading practices and validity. *Educational Measurement: Issues and Practice*, 10(1), 35-36. <https://doi.org/10.1111/j.1745-3992.1991.tb00182.x>
- Brookhart, S. M. (1994). Teachers' grading: Practice and theory. *Applied Measurement in Education*, 7(4), 279-301. Retrieved from <http://eds.a.ebscohost.com.ezproxy.rowan.edu/ehost/pdfviewer/pdfviewer?vid=1&sid=2a525cf7-02a1-4e98-b17d-f05bbae69f64%40sdc-v-sessmgr02>
- Brookhart, S. M. (2004). *Grading*. Upper Saddle River, NJ: Pearson Education Inc.
- Brookhart, S., Stiggins, R., McTighe, J., & William, D. (2019). *The future of assessment practices: Comprehensive and balanced assessment systems*. Paper presented at the Learning Sciences International, College Park, MD. Retrieved from https://www.learningsciences.com/wp/wp-content/uploads/2019/03/LSI_Policy_Paper_3_26_19.pdf

- Brown, A. D., Colville, I., & Pye, A. (2014). Making sense of sensemaking in organization studies. *Perspectives of Organization Studies*, 36(2), 265-277. <https://doi.org/10.1177/0170840614559259>
- Butrymowicz, S. (2017). Most colleges enroll many students who aren't prepared for higher education. Retrieved from <https://hechingerreport.org/colleges-enroll-students-arent-prepared-higher-education/>
- Canady, R. L., & Hotchkiss, P. R. (1989, September). It's good score! Just a bad grade. *Phi Delta Kappa International*, 71(1), 68-71. Retrieved
- Carey, T., & Carifio, J. (2012). The minimum grading controversy: Results of a quantitative study of seven years of grading data from an urban high school. *Educational Researcher*, 41(201), 201-208. <https://doi.org/10.3102/0013189X12453309>
- Carnegie Mellon University. (2019). <https://www.cmu.edu/teaching/assessment/basics/grading-assessment.html>
- Chartier, G. (2003, Spring). Truth-telling, incommensurability, and the ethics of grading. *Brigham Young University Education and Law Journal*, 37, 37-81. Retrieved from <https://digitalcommons.law.byu.edu/elj/vol2003/iss1/3>
- Clark, A. (2016). N.J. revises, renames Common Core academic standards. Retrieved February 15, 2020, from https://www.nj.com/education/2016/05/nj_common_core_standards_christie.html
- Coburn, C. E. (2001, Summer). Collective sensemaking about reading: How teachers mediate reading policy in their professional communities. *Educational Evaluation and Policy Analysis*, 23(2), 145-170. Retrieved from https://search-proquest-com.ezproxy.rowan.edu/docview/197229018?accountid=13605&rfr_id=info%3Axri%2Fsid%3Aprimo
- Coburn, C. E. (2004). Beyond decoupling: rethinking the relationship between the institutional environment and the classroom. *Sociology of Education*, 77(3), 211-244. <https://doi.org/10.1177/003804070407700302>
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Thousand Oaks, CA: Sage.
- Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). Thousand Oaks, California: Sage.

- Cross, L. H., & Frary, R. B. (1999). Hodgepodge grading: Endorsed by students and teachers alike. *Applied Measurement in Education*, 12(1), 53-72.
https://doi.org/10.1207/s15324818ame1201_4
- Cutler, L., & Graham, S. (2008, November). Primary grade writing instruction: A national survey. *Journal of Educational Psychology*, 100(4), 907-919.
<https://doi.org/10.1037/a0012656>
- Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement* (5th ed.). Englewood Cliffs, NJ: Prentice-
- Eccles (Parsons), J., Midgley, C., & Adler, T. F. (1984). Grade-related changes in the school environment: Effects on achievement and motivation. *Advances in Motivation and Achievement*, 3, 283-331. Retrieved from
file:///C:/Users/lanam/Downloads/Grade-Related_Changes_In_The_School_Environment_Ef.pdf
- Ellerbrock, C. R., Denmon, J., Owens, R., & Lindstrom, K. (2015). Fostering a developmentally responsive middle to high school transition: The role of transition supports. *Middle Grades Research Journal*, 10(1), 83–101. Retrieved from <https://search-ebSCOhost-com.ezproxy.rowan.edu/login.aspx?direct=true&db=eue&AN=102933268&site=ehost-live>
- Frisbie, D. A., & Waltman, K. K. (1992). Developing a Personal Grading Plan. *Educational measurement, issues and practice*, 11(3), 35-42.
<https://doi.org/10.1111/j.1745-3992.1992.tb00251.x>
- Graham, S., & Perin, D. (2007a). A meta-analysis of writing instruction for adolescent students. *Journal of educational psychology*, 99(3), 445-476.
<https://doi.org/10.1037/0022-0663.99.3.445>
- Graham, S., & Perin, D. (2007c). *Writing next: Effective strategies to improve writing of adolescent middle and high school*. Retrieved from A report to Carnegie Corporation of New York. Washington, DC: Alliance for Excellent Education: https://www.carnegie.org/media/filer_public/3c/f5/3cf58727-34f4-4140-a014-723a00ac56f7/ccny_report_2007_writing.pdf
- Gronlund, N. E. (1985). *Measurement and evaluation in teaching* (5th ed.). New York, NY: Macmillan.
- Gurung, R. A., & Prieto, L. R. (Eds.). (2009). Teaching diversity through literature. *Getting culture* (pp. 171-179). Sterling, VA: Stylus.

- Guskey, T. R. (1994, October). Making the grade: What benefits students? *Educational Leadership*, 52(2), 14-20. Retrieved from <http://ezproxy.rowan.edu/login?url=https://search.proquest.com/docview/62633613?accountid=13605>
- Guskey, T. R. (2002, April). *Perspectives on grading and reporting: Differences among teachers, students, and parents*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA. Retrieved from <https://files.eric.ed.gov/fulltext/ED464113.pdf>
- Guskey, T. R. (2009, April). *Bound by tradition: Teachers' views of crucial grading and reporting issues*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA. Retrieved from <https://files.eric.ed.gov/fulltext/ED509342.pdf>
- Guskey, T. R., & Bailey, J. M. (2001). *Developing grading and reporting systems for student learning*. Thousand Oaks, California: Corwin Press.
- Guskey, T. R., & Brookhart, S. M. (2019). Introduction. In T. R. Guskey & S. M. Brookhart (Eds.), *What we know about grading: What works, what doesn't, and what's next* (pp. 1-11). Alexandria, VA: ASCD.
- Guskey, T. R., & Brookhart, S. M. (Eds.). (2019). Conclusion: Where do we go from here? *What we know about grading: What works, what doesn't, and what's next* (pp. 214-223). Alexandria, VA: ASCD.
- Hay, P. J., & Macdonald, D. (2008). (Mis)appropriations of criteria and standards-referenced assessments in a performance-based subject. *Assessment in Education: Principles, Policy & Practice*, 15(2), 153-168. <https://doi.org/10.1080/09695940802164184>
- Herr, K., & Anderson, G. L. (2015). *The action research dissertation: A guide for students and faculty* (2nd ed.). Washington DC: Sage.
- Hodder, I. (2012). The interpretation of documents and material culture. In J. Goodwin (Ed.), *Sage Biographical Research* (pp. 171-188). Retrieved from file:///C:/Users/Lana/Downloads/Wk3_Hodder.pdf
- Hopkins, K. D. (1998). *Educational and psychological measurement and evaluation* (eighth ed.). Boston, MA: Allyn & Bacon.
- Krathwohl, D. R., & Smith, N. L. (2005). *How to prepare a dissertation proposal: Suggestions for students in education & the social and behavioral sciences*. Syracuse, NY: Syracuse University Press.

- Kunnath, J. P. (2016). *A critical pedagogy perspective of the impact of school poverty level on the teacher grading decision-making process* (Doctoral dissertation). Retrieved from <http://ezproxy.rowan.edu/login?url=https://search.proquest.com/docview/1762747575?accountid=13605>
- Lindner, K. A. (2017). *Reading workshop and formative assessment: Maximizing quality reading instruction* (Doctoral dissertation, University of Wisconsin-Milwaukee). Retrieved from <https://dc.uwm.edu/cgi/viewcontent.cgi?article=2510&context=etd>
- Link, L. J. (2019). Leadership for grading reform. In T. R. Guskey & S. M. Brookhart (Eds.), *What we know about grading: What works, what doesn't, and what's next* (pp. 157-194). Alexandria, VA: ASCD.
- Liu, X. (2007). *Measuring teachers' perceptions of grading practices: A cross-cultural perspective* (Doctoral dissertation). Retrieved from <https://search-proquest-com.ezproxy.rowan.edu/docview/304862719?pq-origsite=summon>
- Liu, X. (2008, October). *Measuring teachers' perceptions of grading practices: Does school level make a difference?* Paper presented at the Northeastern Educational Research Association Annual Conference, Rocky Hill, CT. Retrieved from https://opencommons.uconn.edu/cgi/viewcontent.cgi?referer=https://www.google.com/&httpsredir=1&article=1011&context=nera_2008
- Llosa, L. (2008, August 28). Building and supporting a validity argument for a standards-based classroom assessment of English proficiency based on teacher judgments. *Educational Measurement Issues and Practice*, 27(3), 32-42. Retrieved from <https://doi-org.ezproxy.rowan.edu/10.1111/j.1745-3992.2008.00126.x>
- McCarthy, S. J., & Ro, Y. S. (2011). Approaches to writing instruction. *Pedagogies: An international journal*, 6(4), 273-295. <https://doi.org/10.1080/1554480X2011.604902>
- McMillan, J. H. (2001, March). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice*, 20(1), 20-32. Retrieved from <https://doi-org.ezproxy.rowan.edu/10.1111/j.1745-3992.2001.tb00055.x>
- McMillan, J. H. (2003, December). Understanding and improving teachers' classroom assessment decision making: Implications for theory and practice. *Educational Measurement: Issues and Practice*, 22(4), 34-43. <https://doi.org/10.1111/j.1745-3992.2003.tb00142.x>

- McMillan, J. H. (2019). Surveys of teachers' grading practices and perceptions. In T. R. Guskey & S. M. Brookhart (Eds.), *What we know about grading: What works, what doesn't, and what's next* (pp. 84-112). Alexandria, VA: ASCD.
- McMillan, J. H., Myran, S., & Workman, D. (2002, March). Elementary teachers' classroom assessment and grading practices. *The Journal of Educational Research*, 95(4), 203-213. <https://doi.org/10.1080/00220670209596593>
- McMillan, J. H., & Nash, S. (2000). Teacher classroom assessment and grading practices decision making. *Paper presented at the Annual Meeting of the National Council on Measurement in Education*. Retrieved from <https://files.eric.ed.gov/fulltext/ED447195.pdf>
- Merriam, S. B. (2001). *Qualitative research and case study applications in education*. San Francisco, CA: Jossey-Bass .
- Messick, S. (1989, March 1). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, 18(2), 5-11. Retrieved from <https://doi.org/10.3102/0013189X018002005>
- Morgan, L. P., & Hertzog, J. C. (2001, March). Designing comprehensive transitions. *Principal Leadership, High School ed.*, 1(7), 10-16. Retrieved from <http://ezproxy.rowan.edu/login?url=https://search-proquest-com.ezproxy.rowan.edu/docview/233325249?accountid=13605>
- Muñoz, M. A., & Guskey, T. R. (2015). Standards-based grading and reporting will improve education: making clear linkages between standards, assessment, grading, and reporting that are concisely reported work for the betterment of ALL students. *Phi Delta Kappan*, 96(7), 64-68. Retrieved from <http://eds.a.ebscohost.com.ezproxy.rowan.edu/ehost/pdfviewer/pdfviewer?vid=1&sid=41c4a2f2-40ba-4ccc-9ead-5e70998f3c0b%40sdc-v-sessmgr05>
- National Center for Education Statistics. (2008). *The nation's report card: Writing 2007*. Retrieved from Washington DC: Institute of Education Sciences, U.S. Department of Education: https://nces.ed.gov/nationsreportcard/pdf/main2007/2008468_1.pdf
- National Center for Education Statistics. (2012). *The nation's report card: Writing 2011*. Retrieved from Washington, DC: Institute of Education Sciences, U.S. Department of Education: <https://nces.ed.gov/nationsreportcard/pdf/main2011/2012470.pdf>
- National Commission on Writing in American Schools and Colleges. (2003). *The neglected "R": The need for a writing revolution*. Retrieved from https://www.nwp.org/cs/public/download/nwp_file/21478/the-neglected-r-college-board-nwp-report.pdf?x-r=pcfile_d

- O'Connor Grading. (2018). <https://www.oconnorgrading.com/>
- O'Connor, K. (1995). Guidelines for grading that support learning and student success. *NASSP Bulletin*, 79(571), 91-101. Retrieved from <https://doi.org/10.1177/019263659507957112>
- Ornstein, A. C. (1994). Grading practices and policies: An overview and some suggestions. *NASSP Bulletin*, 78(561), 55–64. Retrieved from <https://doi.org/10.1177/019263659407856111>
- Patton, M. Q. (2002). *Qualitative research & evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage.
- Persky, H. R., Daane, M. C., & Jin, Y. (2003). *The nation's report card: Writing 2002*. Retrieved from U.S. Department of Education, Institute of Education Sciences. Washington, DC: National Center for Education Statistics: <https://nces.ed.gov/nationsreportcard/pdf/main2002/2003529.pdf>
- Pollington, M. F., Wilcox, B., & Morrison, T. G. (2011, January). Self-perception in writing: The effects of writing workshop and traditional instruction on intermediate grade students. *Reading Psychology*, 22(4), 249-265. <https://doi.org/10.1080/02702710127640>
- Queen, J. A. (2002). *Student transitions from middle to high school: Improving achievement and creating a safer environment*. New York, NY: Routledge.
- Randall, J., & Englehard, G. (2009, Spring). Examining teacher grades using Rasch measurement theory. *Journal of Educational Measurement*, 46(1), 1-18. Retrieved from <https://www-jstor-org.ezproxy.rowan.edu/stable/20461911>
- Rossman, G. B., & Rallis, S. F. (2017). *An introduction to qualitative research: Learning in the field* (4th ed.). Thousand Oaks, Ca.: Sage.
- Rubin, H. J., & Rubin, I. S. (2012). *Qualitative interviewing: The art of hearing data* (3rd ed.). Washington DC: SAGE.
- Rubin, H. J., & Rubin, I. S. (2012). *Qualitative interviewing: The art of hearing data* (3rd ed.). Washington DC: Sage.
- Saldaña, J. (2016). *The Coding Manual for Qualitative Researchers*. Thousand Oaks, CA: Sage.
- Scott, W. R. (2003). *Organizations: Rational, natural, and open systems* (5th Ed. ed.). Upper Saddle River, NJ: Prentice Hall.

- State of New Jersey Department of Education. (1996-2015).
<https://rc.doe.state.nj.us/runreport.aspx?county=25&district=3160&school=057&year=2015-2016>
- State of New Jersey Department of Education. (1996-2017).
<https://www.nj.gov/education/cccs/>
- Stiggins, R. J., Frisbie, D. A., & Griswold, P. A. (1989, June). Inside high school grading practices: Building a research agenda. *Educational Measurement: Issues and Practice*, 8(2), 5-14. <https://doi.org/10.1111/j.1745-3992.1989.tb00315.x>
- Stringer, E. T. (2014). *Action research* (4th ed.). Los Angeles: Sage.
- Sun, Y., & Cheng, L. (2013). Teachers' grading practices: meaning and values assigned. *Assessment in Education: Principles, Policy & Practice*, 21(3), 326-343. Retrieved from <https://doi.org/10.1080/0969594X.2013.768207>
- Teacher Preparation Issues, Docket ID ED-2014-OPE-0057 Department of Education § 205 (2016).
- Teachers College, Columbia University. (2014). Research base underlying the teachers college reading and writing workshop's approach to literacy instruction. Retrieved from <https://readingandwritingproject.org/about/research-base>
- The College Board. (2003). *The neglected "R": The need for a writing revolution*. Retrieved from https://www.nwp.org/cs/public/download/nwp_file/21478/the-neglected-r-college-board-nwp-report.pdf?x-r=pcfile_d
- Toma, J. D. (2006). Approaching rigor in applied qualitative research. In C. Conrad & R. Serlin (Eds.), *Sage handbook for research in education*, pp. 405-423). Retrieved from file:///C:/Users/Lana/Downloads/Wk5_Toma%20(1).pdf
- Troia, G. A., Olinghouse, N. G., Zhang, M., Wilson, J., Stewart, K. A., Mo, Y., & Hawkins, L. (2017, December 29). Content and alignment of state writing standards and assessments as predictors of student writing achievement: An analysis of 2007 national assessment of educational progress data. *Springer*, 31, 835-864. <https://doi.org/10.1007/s11145-017-9816-3>
- U.S. Department of Education. (2017). Developmental education challenges and strategies for reform. Retrieved from <https://www2.ed.gov/about/offices/list/opepd/education-strategies.pdf>
- Von Frank, V. (2009, August/September). Leadership teams create lasting change. *National Staff Development Council*, 13(1), 1-8. Retrieved from file:///C:/Users/Lana/Downloads/Wk6_VonFrank.pdf

- Webster, K. L. (2011). *High school grading practices: Teacher leaders' reflections, insights, and recommendations* (Doctoral dissertation). Retrieved from <http://ezproxy.rowan.edu/login?url=https://search.proquest.com/docview/929134496?accountid=13605>
- Weick, K. E., Sutcliffe, K. M., & Obstfeld, D. (2005, August). Organizing and the process of sensemaking. *Organization Science*, 16(4), 409-421. Retrieved from <http://www.jstor.org/stable/25145979>
- Wieman, C. (2012, Fall). Applying new research to improve science education. *Issues in Science and Technology*, 29(1), 25-32. Retrieved from [file:///C:/Users/Lana/Downloads/out%20\(5\).pdf](file:///C:/Users/Lana/Downloads/out%20(5).pdf)
- Wiley, C. R. (2011). *Profiles of teacher grading practices: Integrating teacher beliefs, course criteria, and student characteristics* (Doctoral dissertation). Retrieved from <http://ezproxy.rowan.edu/login?url=https://search.proquest.com/docview/887719048?accountid=13605>
- Winton, S. (2012, July 24). From zero tolerance to student success in Ontario, Canada. *Educational Policy*, 27(467), 467-498. <http://dx.doi.org/10.1177/0895904812453994>
- Wright, A. (2015). *Defining the other and operationalizing diversity: Lecture 2A EDSU28706* [PowerPoint slides]. Retrieved from https://rowan.instructure.com/courses/1763086/external_tools/146650
- Yesbeck, D. M. (2011). *Grading practices: Teachers' considerations of academic and non-academic factors* (Doctoral dissertation). Retrieved from <http://ezproxy.rowan.edu/login?url=https://search.proquest.com/docview/913076079?accountid=13605>
- Yin, R. K. (2009). *Case study research: Design and methods* (5th ed.). Washington DC: Sage.

Appendix A

Qualitative Informed Consent



Consent to Participate in Research

You are being asked to participate in a research study “Teachers’ Sensemaking: Middle School and High School Language Arts Grading and Assessing Practices for Writing”

Before you agree, the investigator must tell you about

- i. the purposes, procedures, and duration of the research.
- ii. any procedures which are experimental.
- iii. any reasonably foreseeable risks, discomforts, and benefits of the research.
- iv. any potentially beneficial alternative procedures or treatments; and
- v. how confidentiality will be maintained.

Where applicable, the investigator must also tell you about

- i. any available compensation or medical treatment if injury occurs;
- ii. the possibility of unforeseeable risks;
- iii. circumstances when the investigator may halt your participation;
- iv. any added costs to you;
- v. what happens if you decide to stop participating;
- vi. when you will be told about new findings which may affect your willingness to participate; and
- vii. how many people will be in the study.

If you agree to participate, you must be given a signed copy of this document and a written summary of the research.

You may contact Dr. Monica Kerrigan at 856-256-4500 x53658, kerriganm@rowan.edu any time you have questions about the research.

You may contact Dr. Monica Kerrigan at 856-256-4500 x53658, kerriganm@rowan.edu if you have questions about your rights as a research subject or what to do if you are injured.

Your participation in this research is voluntary, and you will not be penalized or lose benefits if you refuse to participate or decide to stop.

Appendix B

Interview Log and Introductory Comments

Date:	Time:
Duration of Interview:	Site:
Participant Gender:	Participant Grade Level:

Introductory Comments

Hello, and thank you for granting me this interview. My name is Lana Cook, and I teach in this district. I'm also a doctoral candidate at Rowan University, and the questions I will ask you today are a part of my dissertation on grading. This interview is a part of the first phase of my research on teacher sensemaking of their writing grading and assessing practices in middle school and high school. The purpose of this interview is to understand how teachers make grading decisions. I want to remind you that your participation is voluntary, and I want to confirm that you received and signed the informed consent before we proceed.

Thank you.

I have several questions to ask you, and I encourage you to be open and honest in your response. I want to assure you that your responses will remain anonymous and confidential and will be used only for the purpose of this independent study. I will record our interview and I will transcribe it using an anonymous code to protect your identity.

Thank you for your participation in this study.

Appendix C

Alignment Matrix

Research Question	Theory	Data Source (*IP=interview protocol; DP=document protocol)	Analysis technique
What contextual factors inform middle school and high school language arts teachers' grading/ assessing practices for writing?	These questions will get at the heart of the “taking cues from the social establishment ” aspect of sensemaking by addressing how teachers grading practices are influenced by contextual factors	IP.3. How does the number of students you teach affect the way you grade their writing? IP.8. What are some grading issues that you may discuss with other teachers?	IP: Hybrid coding: Provisional, Process, and In Vivo. DP: concept coding
What are teachers' values and beliefs about the age group they teach, and learning and motivation in general; and, how do these beliefs influence their writing grading practices (adapted from McMillan, 2019)?	These questions will address another component of sensemaking: interpreting and individually adapting grading practices	IP.9. How did you learn how to grade writing? IP.10. When you look at a _____ grader's essay, what are your expectations? Are your expectations the same as what's on the rubric? IP.12. You teach _____ grade; If a new teacher started teaching here tomorrow, what do you think he or she would need to know about grading this particular age group?	IP: Hybrid coding: Provisional, Process, and In Vivo. DP: concept coding
How do middle school and high school language arts teachers make sense of	These questions will get at the heart of identifying if the grading	IP.4. When you think about grading students' writing, what are the first thoughts that pop into your mind?	IP: Hybrid coding: Provisional, Process, and In Vivo.

their writing grading practices?	paradigms differ between middle school and high school; they will address the idea of coherence of grading practices.	IP.7.How much autonomy do you feel you have with grading students' writing? IP.11. How do you communicate grading expectations prior to assigning a grade?	DP: concept coding
What are teachers' experiences with grading writing in their respective classrooms?	These questions will address a final component of sensemaking: engaging in action. These questions will help isolate the grading practices that are used in each teacher's respective classroom.	IP.6. What writing assessments besides essays do you assess or grade? IP.13. Can you describe a writing assessment you've created and used recently?	IP: Hybrid coding: Provisional, Process, and In Vivo.

Appendix D

Interview Protocol

1. Describe your teaching background.
2. What grade do you teach and how many students do you currently have?
3. How does the number of students you teach affect the way you grade their writing?
4. When you think about grading students' writing, what are the first thoughts that pop into your mind?
5. Can you walk me through your process in grading a batch of essays, from the moment you collect them to the moment you return them? What is the hardest Part of grading students' writing?
6. What writing assessments besides essays do you assess or grade?
7. How much autonomy do you feel you have with grading students' writing?
8. What are some grading issues that you may discuss with other teachers?
9. How did you learn how to grade writing?
10. When you look at a ____ grader's essay, what are your expectations? Are your expectations the same as what's on the rubric?
11. How do you communicate grading expectations prior to assigning grades?
12. You teach ____ grade; If a new teacher started teaching here tomorrow, what do you think he or she would need to know about grading this particular age group?
13. Can you describe a writing assessment you've created and used recently?

14. How do you typically offer feedback to students?
15. What do you view as best practice in grading / assessing students' writing?
16. Is there anything that helps or challenges you with implementing these practices in your classroom?
17. If you overheard two students talking about how you grade/ assess their writing, what do you think they would say?
18. When students move on from your classroom at the end of the year, do you think they can expect the same or similar grading practices from their teacher next year? What may be different?

Appendix E

Document Protocol

1. Who created this grading document and who is expected to adhere to it (is it district policy for teachers/ teacher created for students)? Who, if anyone, contributed to or influenced the creation of this document, besides the classroom teacher?
2. What can I infer is important to the creator of this document? What information is presented, left out, and/ or prioritized on the page, by spacing, order, or otherwise?
3. What is the grading policy that is communicated in this document?