Borderline Grading Decisions: The Factors that Influence the Choices of High School Physical Science Teachers

Kate Roberts
University of Maine, katie.m.hayes@maine.edu

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

Part of the Science and Mathematics Education Commons

Recommended Citation

This Open-Access Thesis is brought to you for free and open access by DigitalCommons@UMaine. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of DigitalCommons@UMaine. For more information, please contact um.library.technical.services@maine.edu.
BORDERLINE GRADING DECISIONS: THE FACTORS THAT INFLUENCE
THE CHOICES OF HIGH SCHOOL PHYSICAL
SCIENCE TEACHERS

By
Kate Roberts
B.S. Northeastern University, 2006
M.S.T. University of Maine, 2009

A DISSERTATION
Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy
(in Education)

The Graduate School
The University of Maine
December 2022

Advisory Committee:

Janet Fairman, Associate Professor of Education, Advisor
Michael Wittmann, Head of Education, American Physical Society
Asli Sezen-Barrie, Associate Professor of Curriculum, Assessment, and Instruction
John Thompson, Professor of Physics
Teachers consider many different kinds of factors in determining student grades. They use a mix of achievement and non-achievement factors in grading decisions, to the criticism of educational measurement experts (Brookhart, 1991; Brookhart, 1993; Frary et al., 1993; Popham, 2009). The factors that influence this decision making are particularly salient when teachers consider borderline grades and make decisions about raising, holding, or lowering marks. The purpose of this phenomenological study was to explore the experiences and self-reported practices of high school physical science teachers’ decision making about borderline grades. Through a series of two semi-structured interviews, twelve participants described the contexts in which they graded and what factors they considered when making final grading decisions. Despite teaching at schools which used different guidelines and policies for grading, the teachers in this study exhibited common patterns of decision making. End-of-term marking periods and the required posting of grades often made teachers consider factors other than academic achievement in their final decisions about student grades. The teachers in this study considered factors like students’ effort or personal extenuating circumstances, but also sought to avoid
perceived negative consequences for students when determining grades. In contrast to prior research, well-documented non-achievement factors: student ability and behavior, were absent in teachers’ decision-making process about borderline grades. Moreover, the teachers in this study expressed concerns about inaccuracy, bias, and subjectivity in their grading and viewed grades as having a margin of error, which appears to be less explored in research on teachers’ grading practices. Teachers utilized certain factors and reasoning consistently at high and low grade borders when deciding whether to increase the final end of term borderline grade. These practices suggest the teachers in this study framed (Tannen, 1978, 1979, 1993; Hammer et al., 2006) borderline grading differently for each border, applied different schemas to different borders, and viewed grading as more than solely the assessment of academic achievement. Previous research has characterized teachers’ overall grading practices are “hodgepodge” and idiosyncratic (Brookhart, 1991; Cizek et al, 1995; Cross & Frary, 1999). The results of this study contradict this characterization and found that teachers were consistent in their decision-making about students’ end-of-term borderline grades along particular borders, regardless of differences in school or district grading guidelines.
DEDICATION

This dissertation is dedicated to my daughter, Evangeline Hayes.
ACKNOWLEDGEMENTS

Thank you to the twelve participants who agreed to be interviewed as part of this study. We were all teaching during an unprecedented time and global pandemic. I cannot thank you enough for your time and willingness to meet with me, share your thoughts and experiences.

I offer my deepest gratitude to my dissertation committee. You have been incredibly supportive. Thank you so much for your feedback, insight, and patience with me along the research journey. This has degree has taken longer than I expected and perhaps you as well. John and Asli, thank you for encouraging me to explore paths I might have overlooked. Beth, thank you for reminding me to be thoughtful about my use of language and choice of words. Michael and Janet, your guidance and support has been invaluable. Michael, you sparked my interest in enthusiasm in educational research. Thank you for guiding me on this journey, despite all the turns, in both research interest and lives. Janet, your guidance and feedback have been essential in this process. Thank you for pushing me and supporting me. I could not have done this without you.

I also want to thank all my fellow single moms who understood and supported my desire to get this degree. Your support was invaluable. You watched Evie for me while I was taking classes. You offered extra-long playdates when I needed to get work done. You even took her trick-or-treating when I could not. To all of you, I am eternally grateful.

Thank you, Mom, for coming in to watch Evie whenever I had evening classes and taking her over weekends so I could work.

Finally, I offer my deepest gratitude to my husband and daughter. David, once we got married, I was able to get so much more work done. Thank you for taking on a second round of fatherhood and believing in me when I didn’t believe in myself. You are a wonderful stepfather
and I could not have done this without your support. Evie, thank you for being patient, understanding, and having a wisdom beyond your years. You have always granted me grace throughout this process and, through your maturity, afforded me the ability to maintain balance. I hope that this degree will serve as an inspiration to you in whatever path you take in life.
# TABLE OF CONTENTS

**DEDICATION**.............................................................................................................iii  
**ACKNOWLEDGEMENTS**..........................................................................................iv  
**LIST OF TABLES**.....................................................................................................xii  
**LIST OF FIGURES**....................................................................................................xiii  

## 1. INTRODUCTION .................................................................................................... 1  
- Research Purpose and Questions ........................................................................... 2  
- Research Positionality and Research Motivation .................................................. 2  
- Significance Statement ............................................................................................ 6  
- Definition of Terms ................................................................................................. 6  
- Organization of the Dissertation ............................................................................ 7  

## 2. REVIEW OF THE LITERATURE .......................................................................... 9  
- Research on Assessment .......................................................................................... 9  
- Assessment Practices of Secondary Science Teachers .......................................... 13  
- Research on Grading ............................................................................................... 13  
  - A focus on grading for academic achievement .................................................... 16  
- Factors that Influence Grading Practices ............................................................... 17  
- Teacher Beliefs about Assessment and Grading ................................................... 17  
- Factors that Influence Teachers’ Grading Practices ............................................. 18  
  - Outside Influences. ............................................................................................... 18  
  - Classroom Influences .......................................................................................... 20
3. METHODOLOGY ..................................................................................................................... 30

Research Design ...................................................................................................................... 30

Participant Sample and Selection Criteria ............................................................................. 32

Participant Recruitment .......................................................................................................... 33

Data Collection .......................................................................................................................... 34

Data Preparation, Coding, and Analysis .................................................................................. 36

Transcription Procedures ......................................................................................................... 36

School, Participant, and Question Summaries .......................................................................... 37

Coding Procedures ................................................................................................................... 37

Thematic coding ....................................................................................................................... 38

Process-Oriented Coding .......................................................................................................... 39

Involved in a Grading Activity ................................................................................................. 40

A Set Pre-Determined Grading Criteria .................................................................................. 40

The Possibility of a New Grading Criteria ................................................................................. 40

A Violation of Expectations and Teacher Reactiveness ............................................................ 40

Trustworthiness and Efforts to Reduce Researcher Bias .......................................................... 41

Participant Recruitment .......................................................................................................... 41
Interviews and Member Checking ................................................................. 42
Fieldnoting and Memoing ............................................................................ 43
Feedback from Others ................................................................................ 43
Summary ........................................................................................................ 43

4. REPORTING OF FINDINGS: GRADING CONTEXTS ........................................ 45

State Context ................................................................................................ 45
Sample Schools ............................................................................................ 46
Science Teacher Participants ....................................................................... 51
Three Approaches to Grading Policy and Practice ..................................... 53
  Proficiency-Based Schools ....................................................................... 55
  Many District Guidelines ......................................................................... 56
  Fewer District Guidelines ....................................................................... 59
Summary ........................................................................................................ 61

5. REPORTING OF FINDINGS: BORDERLINE GRADING FACTORS ............... 62

Grading Practices in Borderline Cases ......................................................... 62
  Quarter, Course and Semester Grades versus Individual Assignment Grades 62
  Considering Consequences for Students ................................................ 64
  Learning over time .................................................................................. 64
  “Bumping Up” but never “Bumping Down” ............................................. 67
The Importance of Student Effort in Deciding Borderline Grades ............ 70
Borders that Warrant Attention .................................................................. 70
  The Pass/Fail Border ............................................................................. 71
  The A/B Border ...................................................................................... 72
Summary ........................................................................................................................................... 77

Decision Making Considerations and Processes ............................................................................ 78

Concerns with Inaccuracy, Bias, and Subjectivity .......................................................................... 79

Inaccuracy ......................................................................................................................................... 79

Where Inaccuracy Appears .............................................................................................................. 80

Inaccuracy in Point Allocation ........................................................................................................ 81

Inaccuracy and Bias in Deciding what to Assess ........................................................................... 82

When Grading is not Representative of Understanding ................................................................. 82

Subjectivity in Grading .................................................................................................................... 82

The Lack of Discipline-Specific Decision-Making when Deciding to Bump Up ...................... 87

Making Decisions during the Course .............................................................................................. 88

Making Decisions for End of Term Grades .................................................................................... 90

Discipline-Specific Decision-Making When Refusing to Bump Up ............................................ 92

Patterns of Behavior ....................................................................................................................... 96

Schema Switching ............................................................................................................................ 97

Summary ......................................................................................................................................... 98

6. DISCUSSION OF FINDINGS AND CONCLUSION ..................................................................... 100

Review of Broad Findings .............................................................................................................. 101

Connections with Prior Research .................................................................................................. 104

Important Borders in Borderline Grading ..................................................................................... 104

Concerns with Error in Grading .................................................................................................... 106

Error Concerns Manifest Differently at Different Borders ......................................................... 108

Increasing Borderline Grades for Non-Achievement Factors ..................................................... 108
The Mixed Influence of Non-Achievement Factors .............................................................. 109
Effort, Behavior, and Ability .............................................................................................. 110
The Non-Uniform Influence of Effort on Decision-Making at Different Borders ..... 113
Ability not cited as a factor ................................................................................................. 114
Less Studied Non-Achievement Factors .......................................................................... 115
Extenuating circumstances as a non-achievement factor .................................................. 115
Avoidance of Perceived Negative Consequences as a Non-achievement Factor ..... 116
Understanding Borderline Grading Practices ................................................................... 117
Decisions about Borderline Grades as Schema Switching ................................................ 119
Content-Focused Schemas ................................................................................................. 121
Alternate Schemas for Grading .......................................................................................... 122
Alternate Schemas of Grading for the Pass/Fail Border ..................................................... 124
Failing Grades within 1-2 Points of Passing ..................................................................... 124
Failing Grades within 5 points of Passing ....................................................................... 125
Different Reasoning along the Same Border .................................................................... 126
Alternate Schemas of Grading for the A/B Border ............................................................. 126
Holding B+ Grades ............................................................................................................ 127
Raising B+ Grades ............................................................................................................ 127
Viewing Borderline Grades as a Set of Contexts ............................................................... 128
Implications and Recommendations .................................................................................. 134
Implications for Future Research ....................................................................................... 134
Implications for Teacher Practice and Professional Development ..................................... 136
Implications for Policy ..................................................................................................... 136
LIST OF TABLES

Table 3.1 Final Codelist........................................................................................................39

Table 4.1 School Sample .........................................................................................................47

Table 4.2 The 7-Point grading scale used by Marsh Valley HS, Woodland HS.........................47

Hilltop Academy, and Riverside Academy .............................................................................47

Table 4.3 The Modified 10-Point grading scale used by Lakeview High School ....................49

Table 4.4 Science Teacher Sample ........................................................................................53
LIST OF FIGURES

Figure 2.1 Initial Conceptual Model of Borderline Grading Decisions........................................ 27

Figure 4.1 Approaches to Grading for the School Sample .............................................................. 54

Figure 6.1 Prominent Non-Achievement Factors in Teachers' Decision Making ......................... 118

Figure 6.2 Revised Conceptual Model of Borderline Grading ...................................................... 121

Figure 6.3 Alternate Schemas and Factors used to Determine Borderline Grades ....................... 123
CHAPTER 1

INTRODUCTION

Grading is fact of the American school system. Not only does it take up a significant amount of teachers’ professional time (Stiggins, 1991), pre-service teacher programs fail to adequately prepare teachers for the demands of classroom assessment and grading (Plake, 1993). Additionally, grades impact students which can affect their schooling experience and their post-secondary prospects and careers.

Grading reforms have been proposed and implemented in the hope of using grades as a mechanism by which to both improve students’ academic achievement and the overall educational system (Brookhart, 1994). The stakes are high and opinions are varied, with some proposing to get rid of grades all together (Kohn, 2011). Despite these reforms, assessment and grading are complicated areas of teacher practice with large implications for students’ learning, development, and potential future opportunities. Additionally, these recommendations have a significant impact on school and potentially state policy.

Grading experts have advised that grades solely represent a measurement of students’ summative academic achievement. However, in practice, grades in the American school system often also incorporate a variety of non-achievement factors. This is particularly prevalent in cases of borderline grades.

The goal of this study is to provide further insight into teachers’ grading practices, specifically those around borderline grades. My hope is that this study will present a more nuanced understanding of teachers’ borderline grading practices in a way that emphasizes teachers’ experiences and voices. An improved understanding of the context in which these
grading decisions take place and the factors that influence teachers’ decisions can help inform recommendations for teacher practice and grading policy.

**Research Purpose and Questions**

The purpose of this phenomenological study was to describe what factors influence how high school physical science teachers determine borderline grades and how these decisions are made. In this study, “grading” is defined as giving a numerical value to an assignment, test, quarter grade, etc. and “borderline grades” refer to numerical grades that are very close to the school district’s official border or demarcation between academic marks. This study was guided by one broad research questions with two sub-questions:

How do teachers experience grading in borderline situations and make decisions about borderline grades?

a. In what grading context and school setting do teachers make these decisions?

b. What factors influence their grading decisions and how?

In the following sections, I will describe my motivations for pursuing these research questions and my positionality as an in-service high school physical science teacher.

**Research Positionality and Research Motivation**

This phenomenological study proposes to explore the experiences of high school physical science teachers’ grading decisions in determining what numerical score to assign borderline student work. Student work can be defined broadly. Work encompasses the whole array of what teachers assign numerical grades to: homework, laboratory reports, worksheets, quizzes, tests, midterms, finals, progress reports, quarter grades, semester grades, and course grades.

The standard approach to explore teacher practices concerning assessment, grading, and borderline grades has been largely quantitative (Brookhart, 1993; McMillan, 2001; Randall &
Engelhard, 2010). Additionally, prior quantitative research on grading often focused on the variance in types of assessments used, categories that compose a grade, grade levels or subjects taught (Cizek et al., 1995; Duncan & Noonan, 2007; Feldman et al., 1998).

The way in which teachers grade and make decisions about grading is not well understood. Although prior literature does provide insight into the larger patterns of teachers’ grading practices, the large scale of these studies fails to address the complexity of grading decision-making teachers engage in on a regular basis. Thus, I chose to approach this study using qualitative methods.

Interpretivist (or phenomenological) qualitative research explores and investigates the lived experiences of individuals. This area of research and lived experience is rooted in my own experience as a high school science teacher. Not only am I the researcher of this study, but I am also a member of the population represented here. In order to better frame this study and understand its purpose, my experiences, positionality, and motivations need to be presented.

I have been a high school science teacher since 2008. I entered into the profession with enthusiasm and determination. I was passionate about working with high school aged children and providing a more student-centered approach than what I had experienced as a student. My teacher education program prepared me for some aspects of my job. Yet, as often is the case, it also left me woefully underprepared with other aspects of my job – grading and assessment, in particular. Teachers spend a significant amount of classroom time engaging in assessment and grading (Stiggins, 1991). As a young teacher, this was certainly my case. I felt like grading was burdensome, at times dreaded, and not very rewarding.

Early on in my career, I realized I needed to trim down my workload and time spent on grading if I wanted teaching to be a sustainable profession. I was fortunate enough to find a
professional development opportunity that focused on assessment and grading. When I was there, I decided that from that point forward, grades in my classes would primarily be a representation of students’ content understanding. I planned out instruction using *Understanding by Design* (Wiggins & McTighe, 2005). I wrote my assessments first and then planned out instructional activities. I stopped grading homework. I stopped grading for participation. In my current grading practices, I grade summative work based solely on what I perceive to be academic mastery. I mainly grade summative assessments in the form of laboratory notebooks and classroom tests.

Despite trying to better adhere to the recommendations of experts, there are flaws in my methods of grading and assessment. Classroom assessments are notoriously lacking via the standards of educational measurement (Popham, 2009). All of the assessments used in my classrooms are teacher-made.

I have no set curriculum to teach nor predetermined way to assess or grade. For different classes and groups of students, I will modify my grading system. For example, in my Advanced Placement classes, I grade formative work because it generally “brings up student grades”, since their summative assessments scores are quite low. Much like my participants in this study, I adjust my methods of grading based on both the classes I teach and the students that are in them.

My system is not perfect, but it is manageable. It is manageable in the context of my school and the way I present it to students and parents. My colleagues have also found a stable equilibrium in their own grading and assessment structure as well. However, each of these systems and decision-making is unique to the individual teacher. Each system of managing a grade book, assessing student knowledge, and decisions about how to assess students reflects the teacher’s choices. Each individual teacher has their own story to tell about the reconciliation of
all these factors and how it manifests in practice. The way I think about assessment and grading is different than that of my colleagues and yet we work in the same content area, in the same school district. I might disagree with my colleagues’ grading systems or methods of assessment, yet I respect their decisions. Ultimately, we have been forced to reconcile multiple factors to make our jobs more manageable.

I handle borderline grades in much the same way as many of my participants. Borderline grades concern me around the end of marking periods, specifically with regards to quarter, semester, and course grades. I inherently assume error in my grading and when examining grades before I post to our school’s online grading platform. I look for grades that are within 0.5 points of the next letter grade. Any grades within that range that would not be bumped up by our grading software are moved to the lowest decimal number score of the next border, e.g., a 92.4 would be raised to a 92.5, which would be rounded in the software to a 93. Unlike my participants, I do not use individual student attributes, like effort, to determine whether to round up. I always round up a grade below any letter grade border, regardless of my perception of students’ non-achievement factors.

Teaching is a balancing act; a difficult one. Grading and assessment are an especially problematic area. Assessment is essential for learning and good assessment vastly improves student learning outcomes (Brown, 2004; Brown, 2005; Stiggins, 2001). Grading, however, seems more complicated. Teacher preparation programs often do not prepare teachers for the grading demands of the classroom (Plake, 1993). Even with professional development, teacher practices around assessment and grading are slow to change (Olsen and Buchanan, 2019). Grading, despite its many useful functions, presents several problems in practice.
Significance Statement

We cannot adequately prepare teachers for something we have a limited understanding of especially regarding teachers’ classroom grading practices. Grading is an aspect of the profession that teachers engage with on a regular basis. Prior research has provided insight into larger trends of teachers’ grading practice, however, a more fine-grained understanding of how teachers approach borderline grading decisions is needed.

If we want to improve education, particularly in the areas of assessment and grading, teachers’ experience and decision-making needs to be better understood in the context in which it takes place. The goal of this study is to understand high school physical science teacher’s experiences determining borderline grades as they perceive it within the structure of their school. In doing so, I hope to give dimension to teachers’ experience and their decision-making surrounding borderline grades. A better understanding of this specific grading case allows for further insight into teacher practice and research and improvement on grading.

Definition of Terms

Grading: The process of determining and placing a numerical value on any piece of student work or assignment and quarter, semester, or course marks.

Numerical Grading: The use of the standard 0-100 scale (regardless of district cutoffs between marks) to determine grades for student work.

Proficiency-Based Grading: The use of the 1-4 proficiency scale to determine grades on student work, evaluate student skills and/or content knowledge, or evaluate student behavior (1 = does not meet proficiency, 2 = partial meets proficiency, 3 = meets proficiency, 4 = exceeds proficiency).
End-of-Term: This is an overarching term that includes final quarter, semester, and course grades.

Borderline: This term refers to assigning a numerical grade that is very close to the district border for academic marks (between an A/B, B/C, C/D, D/F or pass/fail) or for proficiency levels (Exceeds/Meets, Meets/Partially Meets, Partially Meets/Does Not Meet).

Proficiency-Based Schools: These schools use a 1-4 proficiency scale to grade student work and communicate student performance. This grading policy is a result of Maine’s passage of L.D. 1422.

“Many” Guidelines Schools: Schools that have significant district oversight and district-wide policies that influence what and how teachers can grade in their classroom.

“Fewer” Guidelines Schools: Schools that have some district policies in place, but teachers are largely left to decide their own grading practices and classroom policies.

Formative Assessments: Assessments used to inform instruction and give feedback on students’ academic achievement.

Summative Assessments: Assessments used to measure and quantify students’ academic achievement.

**Organization of the Dissertation**

This dissertation is organized into six chapters. The next chapter address relevant literature in the fields of assessment and grading in addition to examining teachers’ grading practices in borderline grading situations. The third chapter describes the methodology used to collect and analyze data. Two findings chapters are then presented. The first findings chapter provides the context in which teachers made their grading decisions. The state and district contexts will be described and three approaches to school grading will be presented. In the
second findings chapter, factors that influenced teachers’ borderline grading decisions will be described. Finally, the sixth chapter presents a discussion of the findings and suggestions for areas of future research as well as implications for policy and practice.
CHAPTER 2

REVIEW OF THE LITERATURE

The following literature review covers a wide range of research focused on grading and assessment. Grading and assessment are complicated areas of research and teacher practice. This chapter is comprised of four parts: (1) research on teachers’ assessment practices (2) general research on teachers’ grading practices (3) beliefs and factors that influence teachers’ assessment and grading practices and (4) research about teacher grading practices around borderline grades.

In the first section, I will review general research on educational assessment. Research in this area relates to teacher preparation for classroom assessment, which includes grading, and provides insight on teachers’ classroom practices. Grading will be discussed generally in the second section and cover teachers’ grading practices and recommendations for teacher practice. In the third section, factors that influence teachers’ grading and assessment practices will be described. Teacher beliefs about grading and assessment are not often delineated in research, so studies describing both will be discussed. Lastly, research specific to teachers’ borderline grading practices will be described.

Research on Assessment

The distinction between assessment, grading, and grades in the literature is vague. A grade, in a sense, is the final, numerical measure of the assessment of academic achievement in the classroom. Although the current study focused on grading, a meta-analysis of the literature suggests that grading and assessment are often intertwined for teachers (McMillan & Workman, 1998). In order to best understand teachers’ grading practices, their assessment practices need to be understood as well.
Assessment has been categorized as either formative or summative. Formative assessment is also referred to as “classroom assessment” or “assessment for learning” (Stiggins, 2002). Formative assessment provides students with feedback about their academic achievement. Formative assessment results allow teachers to adapt instruction or classroom activities to improve or increase student learning. In addition, it also allows students, parents, and other stakeholders to have some insight on students’ learning progress. In a review of research on formative assessment practices, Black and William (1998) found that the use of formative assessment increased students’ overall academic achievement.

Formative assessment can occur in a variety of ways in a classroom setting and can be informal or formal (Coffey et al., 2011; Ruiz-Primo & Furtak, 2007; Furtak; 2009). Informal formative assessment can be as simple as asking students questions in class, having a class discussion, or observing students engaged in a classroom activity. Formal formative assessments can include specific activities such as “predict-observe-explain” (Furtak, 2009), journaling, or more traditionally, quizzes.

In contrast, summative assessment is used as a measurement of student academic achievement after students’ have progressed through a unit of study (Cizek, 2010). Whereas formative assessments are designed to provide feedback on learning, summative assessments theoretically mark the end of a particular “chunk” of learning (e.g., unit, topic, or course). Traditionally, summative assessments often take the form of tests, quizzes, or standardized tests and are administered after students have received feedback (through formative assessment) on their learning. Since summative assessments are measurements of student learning, these are primarily used for grading purposes, whereas formative assessments are not (Guskey, 2010).
It has been estimated teachers spend about 50% of their time on assessment (Stiggins, 1991; Plake, 1993). Multiple researchers have suggested both pre-service and experienced in-service teachers are unprepared to assess students in the classroom (Plake, 1993; Guskey, 2003; Popham, 2009). Therefore, a conflict exists between the extent to which teachers need to assess students and their lack of preparation to do so. The disconnect between classroom demand and lack of preparation has been a focus of research on teachers and often gets included into studies focusing on as “assessment literacy.”

Assessment literacy is defined as the knowledge of the methods and effective use of assessment (Mertler, 2005; Paterno, 2001). However, it has also been described as: “a dynamic, context-dependent social practice” (Willis et al., 2013, p. 2).

Assessment literacy is exceptionally broad. It encompasses highly specific concerns surround educational measurement like validity (content, construct, and criterion) and reliability (Miller, Linn, and Gronlund, 2010). However, assessment literacy also encompasses the “best use” of assessment data. Some argue being assessment literate requires a deep understanding of test construction and evaluation (Popham, 2009), whereas others evaluate assessment literacy on how teachers use the results of best practice assessments to promote student learning (Guskey, 2003).

The former of the two perspectives which focuses on the knowledge of methods dominates much of the work in assessment literacy. This vein of research has been criticized for the focus on (and abundance of) methods of assessment, rather than a detailed description of the process teachers engage in (Coffey et al., 2011).
Prioritizing knowledge about assessment and educational measurement, versus how to best use assessments to inform instruction, can create conflicts in classroom settings and research about these settings. For example, recommended best practice for instruction to promote student learning often includes teacher-made tests and formative assessments (Guskey, 2003). However, teacher-made tests are often criticized because their results cannot be reliably or validly interpreted as a measure of student understanding (Popham, 2009).

Additionally, Brookhart (1993) and Bishop (1992) found that teachers struggled to navigate a related dichotomy. Teachers were required to be judges of student performance, while simultaneously being advocates for their learning. Thus teachers are required to focus on using assessments and grading to promote student learning while also making sure to utilize assessment and grading as “sound” measurements or judgments of students’ learning.

This division results in different focal points for research and classroom practice. Therefore, there is a tension that exists within teachers’ assessment practices and, ultimately, their grading practices: how to create the “best” assessment versus how “best” to use the assessment to promote student learning. Additionally, the tension reflected in recommendations for teacher practice manifests in teachers’ grading and assessment decision. Teachers simultaneously want to promote student understanding through the use of assessments and grading, while also using assessments and grades to ultimately judge students’ content knowledge.

Teachers’ assessment practices are not necessarily their grading practices, which further complicates the study of grading and assessment. Although this study focused on grading specifically, teacher assessment practices are worth discussing, particularly because they have some influence over teachers’ grading practices. Many studies have looked at what factors
influence teacher practice in both grading and assessment. However, the next section will focus on teachers’ grading practices and factors that influence these practices.

**Assessment Practices of Secondary Science Teachers**

In a survey of approximately 100 high school science teachers, Feldman found that teachers rely on tests, quizzes, lab work, class work and homework to assess student work and assign grades (Feldman et al., 1998). In another large quantitative study of 6-12 teachers, McMillan also found that teacher constructed tests were commonly used in determining student grades (2001). Although differences exist among the frequency of types of assessments used by teachers of different disciplines, high school science teachers, like many other teachers, rely on summative, teacher-created, assessments. Projects, oral exams, performance assessments, portfolios and journals are not often used by science teachers to determine classroom grades (Feldman et al., 1998).

**Research on Grading**

Grading can be defined as the “symbols assigned to individual pieces of student work or to composite measures of student performance on student report cards” (Brookhart et al., 2016, p. 806).

Two common types of grading scales are used in the American school system: the more traditional 0-100 numerical scale and the 1-4 proficiency-based scale (Reeves, 2004). Numerical 0-100 scales fall into two commonly used grading sub-scales: a 10-point numerical scale and a 7-point numerical scale. With a 10-point scale, scores of 90-100 are an “A”, 80-89 are a “B”, 70-79 are a “C” and 60-69 are a “D”. Any score below a 59 is considered failing (Link, 2014; Reeves, 2011). The grade ranges for passing are smaller in the 7-point scale; scores below a 69 are considered failing (Link, 2014; Reeves). There is some variability with cut-offs for marks – the
“A” range can vary between either a 93-100 or 94-100 depending on school or district guidelines. This delineation affects subsequent borders as well; for example, a “B” could be considered scores of either an 85-92 or 86-93. With either numerical scale used, teachers can grade based on total points or use weighted categories (Feldman et al., 1998). Significant variations in the determination of student grades using either numerical scale has been documented since early in the 20th century in a study of the grading practices of 140 geometry teachers (Starch & Elliott, 1913).

Both 0-100 numerical scales have been criticized for having a large range of failing marks relative to passing marks (Reeves, 2004). A 0-4 scale has frequently been advocated for as an alternative grading scale that would decrease the failure range and promote more uniform grading practices (Reeves, 2004).

There has been a push in education for grading to solely reflect academic achievement rather than composite measures (Stiggins, 1991; Stiggins, 2001). Standardized testing is widely accepted and used as a method to evaluate and measure students’ academic achievement (Brookhart, et al., 2016). However, classroom assessment scores do not correlate strongly with standardized testing scores (Brookhart, et al., 2016). From an educational measurement perspective, a classroom grade is the assessment of academic achievement in the classroom (Cizek, 1996; Marzano, 2000; Snowman & Biehler, 2003). Theoretically, a grade should focus solely on academic achievement and be determined through reliable and well created assessment tools that meet the requirements of measurement specialists.

This perspective is at odds with teacher practice. Measurement specialists have criticized both pre- and in-service teachers. Many K-12 teacher education programs do not prepare pre-service teachers to grade and assess students (Mertler, 2005; Plake, 1993).
When instruction does occur, formal K-12 teacher training on classroom assessment and measurement frequently omit the classroom context in which they take place. Teachers develop relationships and rapport with students that often impacts their instruction and assessment. Additionally, much of teachers’ time is spent planning instruction rather than on planning assessments (Airasian & Jones, 1993).

In a quantitative study of K-12 teachers, Brookhart (1993) found that assessment is often cyclically intertwined with instruction; each informing the other. Additionally, an in-depth case study of a single teacher suggests teachers may have difficulty scaffolding learning experiences for students and struggle to use assessment to do so (Buck & Trauth-Nare, 2009).

Larger studies of both K-12 and 9-12 teachers indicate that their assessment and grading practices are highly individualized and idiosyncratic, often incorporating factors other than academic achievement (Bowers, 2011; Brookhart, et al., 2016; Cizek, et al., 1995; Frary, 1992). Student learning and academic achievement does appear to be the primary factor that teachers use to determine K-12 student grades (Brookhart, 1993; Guskey & Link, 2019; McMillan, 2001; Randall & Engelhard, 2010). However, teachers often incorporate non-achievement factors, such as effort, behavior, and work habits into grades and these non-academic factors do play a role in the determination of a grade to a varying degree (Brookhart, 1993; McMillan, 2001). Non-achievement factors will be discussed in more detail in subsequent sections of this literature review.

Even when teachers are focused on assessing student knowledge and achievement, there is a discrepancy in what teachers claim to be assessing versus what they actually assess in the classroom (Bol & Strage, 2006). Although this small, 10-participant qualitative study did not
examine classroom grading practices and assessment use, it is worth noting even when teachers aim to grade students’ academic achievement or content knowledge, issues in practice can exist.

**A focus on grading for academic achievement**

As stated above, there is a push in the educational measurement community for grades to represent student achievement and for classroom assignments to assess that achievement. In theory, good practice for classroom teaching involves having both standards and methods of assessing for academic achievement determined before engaging in instruction or assessment (Wiggins & McTighe, 2005). And yet, classroom assessment and grading has been so often characterized as being a “hodgepodge” (Brookhart, 1991) or messy like a “kitchen sink” (Cizek et al., 1995).

Classroom assessment scores do not correlate strongly with standardized testing scores, which measure students’ academic achievement (Brookhart, et al., 2016). Allen (2005), additionally, critiqued the overall validity of classroom grades as measures of students’ academic achievement often finding that teachers grade in similar, invalid ways to how they were graded as students.

There is a discrepancy between the recommended best grading practices and those that teachers employ. Additionally, this discrepancy might arise because of the conflicting assessment roles teachers possess – advocating for students and their learning while simultaneously being required be the judge of student learning. Previous literature reflects the larger phenomenon of teacher practices misaligning with educational recommendations for best practice.
Factors that Influence Grading Practices

Even in the midst of district policies specifying uniform assessment and grading procedures, many teachers’ practices are still highly individualized (Buck et al., 2009). Even with professional development geared towards changing grading approaches, teachers’ grading practices were challenging to change (Olsen & Buchanan, 2019).

The internal beliefs and philosophies teachers have about teaching, learning and the nature of grading and assessment shape their practices. Research shows that teachers share common beliefs about grading and assessment; however, their practices are often highly varied despite these shared beliefs.

Teacher Beliefs about Assessment and Grading

Teachers largely believe that assessments are important and useful for improving student learning (Brown, 2004). In a study of over 500 elementary school teachers, Brown (2004) also found that these conceptions are widespread among teachers regardless of their age and background.

Teachers also want students to do well, which can be heavily influential on grading practices (Cizek et al. 1995, Kunnath, 2017; McMillan & Nash, 2000; McMillan et al., 2002; McMillan, 2003; Tierney, 2015). This often results in teachers incorporating more non-achievement factors (or academic enablers), like effort, into their grading practices in order to increase student grades as a means of supporting student success.

Grades can also serve as a way to motivate students and keep them involved in the “learning process” (Brookhart, 2004; McMillan & Nash, 2000). Grading scales and zeros are often used in teachers’ grading practices to encourage certain types of student behavior.
Lastly, there is a moral aspect of grading. Grading can be used to guide classroom or course content, but also students’ moral development (Zoekler, 2007). Additionally, teachers’ own moral concerns can cause them to adjust a grade (Tierney, 2015). Teachers want their assessment and grading practices to be fair and often take individual student attributes into account (Airasian & Jones, 1993; Brookhart, 1992). Tierney (2015) found that teachers would modify grades based on compassion, wanting to provide students with opportunities or wanting to teach them life lessons.

Factors that Influence Teachers’ Grading Practices

Various factors can shape teachers’ assessment and grading practices (McMillan & Workman, 2002; Ogan-Bekiroglu, 2009). These influences can be divided into two categories: influences outside the classroom and influences inside the classroom. Influences inside the classroom can include subject taught, class size, grade level, or students’ attributes. Outside influences would encompass school and parental expectations, district policy, or state policy and standardized testing.

Outside Influences. McMillan and Nash (2000) define outside or external factors as those that originate outside of the classroom. Influences outside the classroom can include a variety of factors, like school and parental expectations, school size, school or district policies, and state policy and standardized testing requirements. In an interview study of 24 elementary and high school English and math teachers in Virginia, McMillan and Nash (2000) identified three external influences: high-stakes standardized testing, district policies, and parents. However, these had varied and sometimes little effect on teachers’ assessment and grading practices.
Teachers in this study would modify their assessments to better align and represent items from the state’s standardized test, despite having objections to the use of multiple-choice questions and the ability of those types of questions to provide insight on student learning or guide instruction. Despite having some effect on assessment practices, standardized testing had little effect on teachers’ grading practices.

McMillan and Nash (2000) found that district policies had little to no effect on teachers’ grading and assessment practices, with some teachers ignoring district guidelines altogether. Individual teacher preferences about grading and assessment influenced practices far more than school policy, a finding supported by Cizek et al. (1995).

The influence of parents affected teachers’ grading practices far more than their assessment practices or the types or nature of the assessment used in their classroom (McMillan & Nash, 2000). Teachers primarily focused on having a grading system to minimize conflict with parents concerning grades.

Other studies have been done that look at the influence of school setting or environment on teachers’ grading and assessment practices. School size and school setting (urban, suburban, and rural) have been found to have no effect on teachers’ grading practices (Duncan & Noonan, 2007; Feldman et al., 1998).

Teachers often are managing multiple facets of teaching: simultaneously being cognizant of external responsibilities and classroom factors (Airasian & Jones, 1993). External or non-classroom influences appear to have less impact on teacher practices than internal (or classroom) influences. In the next sections, classroom influences will be discussed in addition to non-achievement factors.
Classroom Influences. Studies by Duncan and Noonan (2007) and Feldman et al. (1998) have examined the effect of classroom learning factors on the grading practices of 9-12 science teachers. Classroom learning factors included elements such as class size, subject taught, and school size. Feldman et al. (1998) surveyed 91 high school science teachers and found that teachers’ years of experience and gender had no statistically significant effect on their assessment and grading practices. Additionally, in a larger survey of 513 high school science teachers, Duncan and Noonan (2007) found that class size had no effect on assessment and grading practices either.

However, both studies found that the type of science taught was the only statistically significant factor that affected teachers’ assessment practices. Teachers in different science subject areas tended to use different types of assessments which resulting in different method of determining course grades. However, despite having statistical significance, the effect of subject area only mildly affected teacher practices.

Larger scale, subject area differences in assessment and grading practices have been found across core subjects (e.g. – English, math, science, social studies). McMillan (2001) found that in a study of approximately 1,500 grade 6-12 teachers of different disciplines used different factors in determining grades, with some certain subjects (English and social studies) using more non-achievement factors than others (math). Teachers of certain disciplines, performing Arts and English teachers tended to incorporate “academic enabling” factors (process, effort, etc.) more often into assessment and grading than math and science teachers. Additionally, like the studies on grade 9-12 science teachers, the types of assessments used by teachers and the relative grade weighting have been shown to be dependent on subject area taught (McLean, 2018; McMillan, 2001).
Two studies have shown that differences in teacher grading and assessment also occur between middle and high school teachers (Cizek et al., 1995; McMillan, 2001). Both of these studies were large quantitative surveys of K-12 and 6-12, respectively. Differences between types of assessment items used and relative weighting of academic and non-academic were found with middle school teachers tending to weight non-achievement factors more heavily. Similar differences have been found between elementary and middle school teachers (Randall & Engelhard, 2009). Elementary teachers incorporated more non-achievement factors into their grading which resulted in higher grades than with middle school.

**Non-Achievement Factors.** It is well documented that teachers often use a mix of achievement and non-achievement factors when determining grades (Brookhart, 1993). Students’ academic achievement can be defined as the acquisition of course content and skills and the measurement thereof, using reliable classroom assessment. Non-achievement factors are not related to a student’s academic achievement. Effort and perceived ability are two commonly cited non-achievement factors (Frary et al., 1993; Brookhart, 1994). Student behavior and work habits have also been used by teachers in their decision making (Guskey, 2011). In large scale surveys of teachers, class participation was found to be an additional non-achievement factor used in grading as well (Cross & Frary, 1999; McMillan & Nash, 2000). However, despite there being several different non-achievement factors teachers can and do consider, effort ability, and behavior tend to be the most influential in practice and most prominent in the literature (Randall & Engelhard, 2010).

Randall and Engelhard (2010) provided approximately 500 K-12 teachers with hypothetical students with an academic achievement-based grade and various non-achievement factors to explore decision making. They provided teachers with descriptions of students’ non-
academic factors (effort, ability, and behavior) in addition to information about their academic achievement K-12 teachers often factored these non-achievement attributes into their final decisions about report grades. Given different scenarios, teachers frequently utilized these non-achievement factors when grades were along the border for a mark.

**Borderline Grades**

“Borderline” refers to numerical or proficiency grades that are very close to the border of district-specified marks. As previously established, throughout all their grading, teachers (to the criticism of the educational measurement community) often grade for non-achievement factors (Brookhart et al., 2016; Duncan & Noonan, 2007; McMillan & Nash, 2000; Randall & Engelhard, 2010; Sun & Cheng, 2014). When teachers engage in making decisions about borderline grades, the difference between achievement and non-achievement classroom factors appears to be more prominent in their reasoning (Randall & Engelhard, 2010; Sun & Cheng, 2014). Teachers used characteristics like effort, behavior, and ability as the primary non-achievement factors in their decisions making (McMillan et al., 2002; Randall & Engelhard, 2010).

Another interesting factor is that teachers weigh achievement and non-achievement factors differently for different students. This is particularly salient with students who are considered borderline cases (e.g. – a 92 in the 7-pt grading system is a B+, but is only one point away from being an A-). Randall and Engelhard (2010) surveyed over 500 K-12 public school teachers and specifically presented them with scenarios of borderline students. In each scenario, student ability, achievement, effort, and behavior were described. All four of these factors were found to have a four-way statistically significant interaction with the final grades these hypothetical students were assigned. Unsurprisingly, students who exhibited high effort and
excellent behavior often were afforded an increase in their borderline course grade. This is in contrast to students displaying low effort and poor behavior who did have their grade raised, or in rare cases, sometimes had their grade lowered.

Brookhart (1993) found a similar pattern in a survey of 84 grade K-12 teachers. Teachers would raise grades for students who exhibited more desirable non-achievement factors with the largest effect being on the lowest grading border – a pattern also seen by Randall and Engelhard (2010). Brookhart summarized:

There is a double standard of just desserts: An average student gets ‘what she earns’, while a below-average students gets a ‘break’ if there is any way to justify it. The difference is in how the teacher perceives the student and reflects the teacher’s advocacy function. (p. 140)

It appears that the discrepancies that appear throughout teachers’ grading and assessment practices, often characterized as idiosyncratic and highly individualized, are more pronounced in borderline grading situations. Furthermore, many studies on grading have been large quantitative studies that focus on determining on variation in teacher practices and influential factors rather than more deeply examining the process by which teachers make their decisions. Different decisions can be the result of different perceptions of a situation. Given prior research on grading, teachers often make different grading decisions despite recommendations to only grade for content knowledge and academic achievement. The different decisions teachers make at different times warrants discussion of the psychological concept of “frames”, which will be discussed in the next section.
Frames and Framing

A “frame,” and later the process of “framing,” as conceptualized in this study is largely informed by the linguistics work of Deborah Tannen (1978, 1979, 1993). Tannen investigated how people used language to convey their expectations about an activity or situation. Drawing on this framework, Hammer et al. (2006) define a frame this way: “By a ‘frame’ we mean, phenomenologically, a set of expectations an individual has about the situation in which she finds herself that affect what she notices and how she thinks to act” (p. 89). In a given frame, an individual notices particular features of the situation, utilizes specific understandings, and engages in different behaviors. Framing consists of determining what frame to apply to a given situation, noticing some things preferentially, and then utilizing different facets of reasoning or behavior. Different framing results in different patterns of behavior.

Framing, Resources, and Schemas

Patterns of framing and consistent resources utilization are similar to the idea of schemas proposed by Piaget (1971). Schemas were first introduced by Piaget in his work studying the cognitive development of children. Schemas are patterns of cognition developed from past experiences. When an individual frames a situation, they draw on set schemas and resources to utilize in understanding or navigating that situation.

Extensive research has been done on epistemic framing and resource use in student learning (Hammer et al., 2006; Hammer & Elby, 2010; Tuminaro & Redish, 2007). The phenomenon of framing is epistemic in nature because it deals with the nature of knowledge, specifically the processes by which knowledge is acquired or restructured. Grading is not epistemic in nature and the current study focused on teachers’ self-reported experiences rather than how teachers acquire or modify their knowledge about grading. However, the framework
Hammer and Elby (2010) describe is applicable to how teachers describe their experiences and decision-making about borderline grades. They describe an epistemological frame as a “locally coherent activation of a network of resources” (p. 409). Resources are fine-grained facets of knowledge that a student (or person) possesses but can be contextually activated. When resources are activated in a frame they can appear as a “belief.” Hammer and Elby (2010) also described frames as being consistently stable in their context when they have been used multiple times (and gain structural stability). Additionally, they found that framing can occur deliberately, when an individual strives to maintain a “consistent stance” towards an activity. Teachers use certain elements of knowledge when determining grades, yet the process of grading is not a knowledge-building or knowledge-validating process. However, the mechanism by which situations are framed, what details are perceived as relevant, and which facets of knowledge are used, might provide insight into the varied practices of teacher grading. Frames and schemas provided a useful lens, allowing for a deeper interpretation and understanding of the self-reported borderline grading practices of teachers. The concept of framing was used in later stages of this study rather than in the development of a conceptual framework, which is discussed in the next section.

**Initial Conceptual Model**

Several key concepts from the literature informed my initial conceptual model for this study which ultimately guided my study design, development of interview protocols, and my preliminary analysis. It has been well-documented that teachers use non-academic factors to determine grades, which drew my focus to the multitude of factors they might consider. I relied heavily on McMillan’s work that described the various factors that affected K-12 teachers’ assessment decision-making rationale (McMillan, 2003). McMillan described three main
influences on teachers’ decision-making rationale: (1) teachers’ knowledge, beliefs, and values, (2) external factors (e.g., state testing or district policy) and (3) classroom realities. Rather than use these categories as presented, I conceptualized them differently. I chose to view teachers’ knowledge and beliefs about assessment simply as potential factors that could influence borderline grading decisions and did not differentiate between knowledge, beliefs, values or expectations. Although I use the term “assessment belief or practice,” this is a very general reference.

The concepts of assessment and grading are interwoven for teachers, and beliefs about assessment or factors that influence assessment might possibly influence grading as well. For example, teachers’ desire to promote student learning, despite being characterized as a belief about assessment, might also be influential on their grading practices. Teachers’ grading practices might also be influenced by their knowledge of assessment and best practices. Additionally, the tension between assessing to promote student learning, but also measure student learning might influence teachers’ grading practices and decision-making about borderline grades.

I also used McMillan’s categorization of “external influences” and “classroom realities” (McMillan, 2003) in my conceptual model, albeit slightly modified. I viewed external influences as outside factors and included many of the same elements McMillan did: district expectations, school policy, administration and parents. However, in contrast, I considered classroom realities as having two separate components: general classroom factors (i.e., class size, subject taught, or academic level) and student factors or attributes. “Student factors” encompasses a wide range of students’ non-academic attributes such as effort, behavior, and ability, which have been well documented in the literature (Frary et al., 1993; Brookhart, 1994, Guskey, 2011; Randall &
Engelhard, 2010). Furthermore, I considered outside and classroom factors to be essential influences on the larger context in which teachers made their grading decisions, whereas assessment beliefs or student factors were not. Figure 2.1 depicts my initial conceptualization of teachers’ borderline grading decisions as I designed and implemented this study.

Figure 2.1 Initial Conceptual Model of Borderline Grading Decisions

Because multiple factors and influences have been found to affect teachers’ grading practices, I anticipated certain factors may be more influential at times than others. I hypothesized that teachers would only be able to pay attention to a limited number of factors given the large quantity they are exposed to. Additionally, borderline grading decisions might warrant different considerations for different borders and certain grading decisions could be differently framed than others.
Although I had identified factors and contexts from prior research that had been found to influence teachers’ general grading practices, I anticipated that borderline grading decision-making could be different. I remained open to uncovering new factors or influences in teachers’ decision-making and representing the self-described experiences and practices of my participants. This initial conceptual model guided my beginning stages of inquiry into this study.

**Summary**

Extensive research has been done surrounding teachers’ grading and assessment practices. Teachers need to understand how to create useful assessments to reliably measure student learning, but also need to be able to use the results of those assessments to promote student learning. Teachers often try to assess and then grade students based primarily on their academic achievement; however, there is significant variation in how this manifests in teachers’ classroom practice. Teachers’ assessment and grading practices have frequently been criticized as being highly variable and inconsistent.

Although the recommended best practice is grading for academic achievement, teachers often incorporate non-achievement or “academic enabling” factors to some degree in their grading. Additionally, the way in which both achievement and non-achievement factors affect grading and grade calculations vary widely across teachers. Students’ academic effort, behavior, and teachers’ perception of student ability frequently affect their grading decisions.

Many of the studies presented in this chapter have been large quantitative studies. Even in the more descriptive studies, large numbers of teachers have surveyed, and teacher responses have been placed into pre-existing categories. Not much research has been done to explore *how* teachers apply what they know and believe to assessment and grading, or that uses a mix of different research methods. Less still has been done on how teachers *experience* this process. The
experience of teachers in the classroom is a valuable area of research. Airasian and Jones (1993) state: “As they are caught up in the many demands of the classroom, teachers can rarely solve problems or reach decisions in general or based solely on theoretical principles” (p.244). The study presented in the next few chapters aims to better explore and understand the grading contexts and factors influence secondary physical science teachers’ decision-making about borderline grades. The next chapter describes the research methods used for this study.
CHAPTER 3

METHODOLOGY

The purpose of this phenomenological study was to describe what factors influence how high school physical science teachers determine borderline grades and how these decisions are made. This research was guided by one broad research question with two sub-questions:

How do teachers experience grading in borderline situations and make decisions about borderline grades?

a. In what type of context and school setting do teachers make these decisions?

b. What factors influence their decisions and how?

This chapter will outline the research design, interview development, participant recruitment and analysis methods for this study.

Research Design

This research focused on how teachers make decisions about and experience borderline grades for students. Given the focus on teachers’ individual, lived experiences, a qualitative research approach was used. Denzin and Lincoln (2011) state: “…qualitative research involves an interpretive, naturalistic approach to the world. This means that qualitative researchers study things in their natural settings, attempting to make sense of, or interpret, phenomena in terms of the meanings people bring to them.” My goals were to discover factors teachers identified as meaningful and then to further explore how these factors influenced their decisions about grading. Additionally, these goals were process-oriented, in that I was interested in how decisions were made rather than what decisions were made.
Maxwell (2013) states that a strength of qualitative research is this focus on process theory. Due to the nature of the research questions and purpose, qualitative inquiry provided a valuable (and necessary) approach.

Another value in approaching the research questions qualitatively was that the voices of participants would be prominent in the findings (Creswell, 2013). This was of particular importance to me as I did not want to engage in a research study solely for the purpose of earning a degree; rather, I felt I could use this research as a platform for teacher voices. As a practicing teacher, I deeply value the voices and lived experiences of the teachers who agreed to be part of this study. Qualitative inquiry allowed me to honor the work of these teachers and explore their diverse experience surrounding grading and assessment.

A phenomenological approach for the research methodology best fit the purpose and research questions of this study. An interpretive, phenomenological method of qualitative research, as described by Creswell (2013), has the goal of describing and exploring the experiences of a group of individuals with some specific phenomenon. The goal of this research was to describe and better understand the borderline grading experiences of high school physical science teachers. This study follows an interpretive approach with a goal of “understanding phenomena in their own right (rather than from some outside perspective)” (Elliott & Timulak, 2005, p. 147).

Qualitative data can be collected in a variety of ways. I chose to collect data through interviews. Again, it was important that these lived experiences be described by the participants themselves, as practicing teachers, as opposed to collecting observational data or school documents. Participants’ recollection of the events, perception, and interpretation are essential to understanding their experience. Qualitative interviews also provided an opportunity to be
responsive to the conversation and participants’ responses and to obtain rich descriptive data. Through the use of open-ended questions, coupled with clarification and follow-up questions, I was able probe participants’ comments in more depth, which would not be possible with other methods, such as surveys. Additionally, it is important to note that this study was conducted during a pandemic and other methods of qualitative data collection (e.g., school visits or observations) were not feasible.

**Participant Sample and Selection Criteria**

To be able to address the research questions, teachers needed to fulfill certain criteria for selection to participate in this study. The focus of this study was on the grading practices of secondary physical science teachers in the state of Maine. Therefore, a purposeful sampling method was used to recruit and select teachers for this study. Prospective participants had to be: (1) located in the state of Maine, (2) teaching at the high school (grades 9-12) level, (3) and teaching physical science classes.

It is important to note that because Maine is a rural state, teachers often have two or more teaching licenses, and teachers often teach more than one subject. In science, it is not uncommon for a secondary science teacher to have dual certification in grades 7-12 in both life science and physical science. Additionally, teachers can often have certifications in other disciplines as well (e.g., mathematics). However, the focus of this study was on physical science only. Even if selected teachers did teach other subjects, only their experience teaching and grading in physical science classes was discussed in the interviews.

The last criteria used to determine participant eligibility was their experience. I purposefully sought out more experienced teachers – specifically, those with four or more years of teaching experience at their current schools. More experienced teachers would be well
acclimated to and aware of their school or district’s grading policies and better able to describe the context in which they made their decisions. Moreover, with increased experience and possibly tenure (if available in their school districts), teachers might feel more freely able to discuss their experiences and views, including negative experiences or influences. In contrast, a new, inexperienced teacher might not be familiar with the grading policies or comfortable enough with the process of determining borderline grades, and might also be less able or comfortable to articulate their experiences and decision-making processes.

The participants in this study were purposefully selected using the four criteria described above: they were experienced teachers, they taught physical science, they taught at the high school level, and they taught at schools in Maine.

**Participant Recruitment**

Convenience sampling (Miles et al., 2020) was used to select teachers from various high schools in Maine, after determining whether teachers met the requirements to participate. I did not seek out certain types of schools (e.g., public or private). Rather, I was focused on getting participation from teachers that met these requirements, especially given this study was being performed during a global pandemic.

To recruit participants, I sent out recruitment emails to teachers throughout the state. The email addresses of these teachers were publicly available on school websites. By starting with the high school websites, I was able to see which teachers met the two initial criteria for the study: they taught at the high school level and taught physical science specifically. In the recruitment email (see Appendix A), I additionally requested that prospective participants have more than four years of teaching experience and/or have tenure at their school.
In my initial round of recruitment emails, I sent out twelve emails to seven different schools, purposefully targeting physical science teachers. I only emailed participants that were in more populated areas of Maine. In northern Maine, a more rural part of the state, many teachers teach multiple subjects, rather than only high school physical science (or only science). Through the school and individual teacher websites coupled with a general familiarity with the teaching community, I was able to identify potential participants. These initial participants were from larger schools to increase the likelihood that they only taught physics or physical science.

The first recruitment email sent out in January 2021 enlisted a few initial participants. A second request for participation was sent in February to the unresponsive initial participants (the same initial twelve). This second request gathered more participants, bringing the initial total to seven teachers who agreed to participate. The remaining study participants were recruited using a snowball technique (Seidman, 2019). During the first round of interviews, I asked participants about the potential participation and recruitment of other high school physical science teachers. This enabled me, in some cases, to have more than one participant from the seven different schools. These schools varied in total student enrollment and geographical location. Additionally, the schools in this study were private and public, and represented the socioeconomic diversity of the state. Each school, its demographics, and its approach to grading are described in more detail in the following chapter.

**Data Collection**

Two rounds of interviews were conducted remotely with each teacher individually by Zoom. The first interview during January and February of 2021 explored teachers’ perceptions about the school or district grading policies and teachers’ practices around grading and assessment generally (see Appendix C). The second interview during May and June of 2021
focused specifically on teachers’ practices, experiences and factors considered related to borderline student grades.

Two interview protocols were created and used in this study. Both received approval from the internal review board (IRB) at the University of Maine. The first interview protocol was developed with different, initial research questions that focused on assessment and grading generally, rather than targeting teacher experiences around borderline grades.

Based on the first interviews, I was able to identify an area in teachers’ grading experiences to further explore. As stated previously, I adamantly wanted this research to emphasize teacher experiences and voices. Although I did not employ collaborative research methodology (Penuel et al., 2020), I used common participant experiences, concerns, and rich areas of discussion to determine a specific area in grading and assessment to focus on. I also shared preliminary samples of the interview data and findings with my thesis committee for discussion and validation of the decision to adjust my research focus and revisit the study’s broad research questions. Subsequently, I decided to focus on borderline grades. The concept of “borderline” drew the focus to specific grades: those on the border between a mark or proficiency level. Additionally, I chose to focus on grading rather than assessment. Grading seemed to be a more commonly represented experience in the first interviews and is a more specific in nature than the concept of assessment.

The interview protocol focused on teachers’ experiences in borderline grading situations and how decisions were made surrounding these grades. This protocol can be found in Appendix D. The second interview protocol consisted of ten interview questions with follow-ups as needed. Two to three specific questions were added to further explore areas of participant
experience based on the first interview. These questions also added an additional layer of member checking to make sure my initial impressions of the first interview data were accurate.

**Data Preparation, Coding, and Analysis**

In this section, the analytic steps taken with the interview data will be described. These include: how the data were transcribed, initial analysis prior to coding, coding procedures, memoing, and fieldnotes.

Prior to coding, several steps were taken to organize and better understand the data. The first step was accurately transcribing and listening to the data to characterize initial impressions. Subsequently, written summaries of the school contexts, participants, and interview questions were developed as descriptive fieldnotes to better understand the data and begin to think about future coding and interpretation.

**Transcription Procedures**

Interviews were conducted remotely over Zoom and both audio and video data were recorded. The video data were not used in the analysis. The audio portion of the recording was then transcribed electronically using transcription software.

After the initial transcription, I went through each transcript twice to correct transcripts for greater accuracy using the audio recordings of the interviews. At this stage of the research, it was unclear as to whether discourse analysis would be used as a future tool for analysis or not. Additionally, as previously mentioned, I wanted to provide an authentic voice for my participants and wanted the transcript to accurately represent what was said.

I edited the initial transcripts for word misidentification, punctuation, emphasis, etc. The second pass focused more on correct speaker identification, wording, pauses in speech, major changes in speech volume or tone, and false starts. During this time, as well as through the
research process, I engaged in descriptive and analytical memoing as described by Saldaña (2016) and Miles et al. (2020). I wrote memos on participant demographic information and how they graded in their classroom: what, when, and how they graded assignments, how they calculated final grades, and the grading structures they operated within. Following this descriptive analysis, I wrote memos on my initial thoughts about how teachers experienced grading and made decisions. I also made note of any participant statements I found particularly interesting. Later, I also wrote about potential analytical coding labels and interesting features in the transcript.

**School, Participant, and Question Summaries**

To better orient to the data, descriptive summaries were written about the seven schools in which participants taught. This information included: the type of grading system used, the grading software used by the school, and policies around grading and assessment. Separate written summaries about each teacher were also developed. These teacher summaries included: subjects taught, years of teaching experience, opinions and beliefs about grading, and self-reported practices around grading, specifically for borderline grades. Lastly, to gain an understanding of commonalities among teachers, I created interview question summaries. In these summaries I would select quotes from each participants’ response to the question and write descriptively about what they said in response to the interview questions, often connecting it to statements said earlier or later in the interviews.

**Coding Procedures**

Following transcription, note taking and writing, a preliminary codelist was developed based on my initial impressions of the data, participant and question summaries, and previous research findings on this topic. My approach to coding was therefore both deductive and
inductive, but more heavily weighted towards an inductive approach, from my close reading of all the interview data. However, based on my review of previous research on teacher practices surrounding borderline grades, some additional coding categories were also created to see how the experiences of this study’s participants might compare to prior research findings.

**Thematic coding.** The actual process of coding included the development of the provisional codelist and then a process of inductive open coding of the interviews (Miles et al., 2020). Coding was conducted using the qualitative data analysis software, ATLAS.ti. As the codelist developed through the coding process, I shared both the codelist and excerpts from that data with my committee members to both validate the coding as well as further refine the final codelist (Table 3.1).

Table 3.1 shows the codelist with both major categories and subcategories. Five initial categories were developed with ten *a priori* codes for the provisional codelist. These were based on initial readings of the data and prior research. The codes shown in italics in the table are the additional codes created inductively through the open-coding process.
Table 3.1: Final Codelist

<table>
<thead>
<tr>
<th>Coding Group</th>
<th>Individual Codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Achievement Factors</td>
<td>Effort, Behavior, Ability, <em>Extenuating circumstances</em></td>
</tr>
<tr>
<td>Teacher Perception of Student Needs</td>
<td>Motivation, Reward/Recognition for effort, <em>Safe place to make mistakes</em></td>
</tr>
<tr>
<td>Time of Year</td>
<td>Motivation early for effort, Easing up later in the year, <em>Growth</em></td>
</tr>
<tr>
<td>Teacher Perception of Stake Level</td>
<td>High, Low</td>
</tr>
<tr>
<td>District Expectations</td>
<td>District Expectations</td>
</tr>
<tr>
<td><em>Flexibility in Practices</em></td>
<td></td>
</tr>
<tr>
<td>More Flexible</td>
<td>Conceding, Adapting</td>
</tr>
<tr>
<td>Less flexible</td>
<td>Taking a stand</td>
</tr>
<tr>
<td>“Grading is Imperfect”</td>
<td>Inaccuracy, Subjectivity</td>
</tr>
</tbody>
</table>

In total, after the open coding, there were eight code categories and 18 individual codes. These codes were then merged into larger themes (Miles et al., 2020), which are discussed in Chapter 4.

**Process-Oriented Coding.** Following this thematic coding, I developed an interpretive, process-oriented coding method that focused on how teachers said they modified their approach to grading based on certain factors they considered; specifically, for determining end of term borderline grades. I categorized teachers’ self-reported behaviors and decision making into a series of four steps or criteria. To identify instances where teachers were changing their approach to grading, these four criteria were examined: teachers needed to be involved in a grading activity, a predetermined method of grading existed, an alternate method of (or schema) for grading was referenced (directly or indirectly) and switching to this alternate schema was reactive on the part of the teacher. Each of these will be discussed in more detail in Chapter 5.
**Involved in a Grading Activity.** To meet this baseline coding requirement, the teacher needed to be describing an experience where they were grading some set of student work. In the interview data, this occurred when the teacher was describing how they determined and assigned a grade to a piece of student work (e.g., worksheet, presentation, test, homework, quarter grade, semester grade, or end of year grade).

**A Set Pre-Determined Grading Criteria.** When describing a grading experience, I looked for teacher references to set criteria or standards they planned to utilize (a rubric, an allocated number of points per question or sub-question, a determined weighted percentage by which a grade will be calculated). The method of grade calculation could be directly stated or implied. For example, if a teacher referenced a rubric they are using to grade student projects, I would be a direct statement of a predetermined schema for grading. Alternatively, during conversation a teacher could question the number of points a student was losing on their test. This statement implies that the teacher had a set point allocation for the test and individual test questions.

**The Possibility of a New Grading Criteria.** While describing their experience, the third criteria for coding occurred when the teacher referenced or acknowledge the ability to grade via a different set of criteria. This could be the referenced directly or indirectly and described as conscious or unconscious choice by the teacher. I coded freely for this criterion: the teacher did not need to describe engaging in this behavior, they only needed to acknowledge it as a possibility.

**A Violation of Expectations and Teacher Reactiveness.** The last criteria for this process-focused code was that the self-described decision or consideration of changing grading, was reactive to a particular situation and context dependent. I categorized this reactiveness in a series of three parts: (a) the teacher had expectations surrounding the particular grading experience
(e.g. – individual student performance, final calculated scores, or student effort), (b) those expectations were violated in the given situational context, (c) the teacher then negotiated the response to the violation of expectations and the teacher is able to move to a different criteria/schema for making grading decisions or at least recognize that this shift is an available option.

These four criteria established a method for coding teachers’ decision making and grading practices. This process-oriented coding approach helped support the development of emergent thematic codes and deepen understanding about teachers’ decision-making process. The findings from this analysis are discussed in Chapter 5.

**Trustworthiness and Efforts to Reduce Researcher Bias**

Throughout this research, multiple steps were taken to minimize potential researcher bias or influence. Prior to conducting this study, the recruitment and selection of participants was designed to minimize my own bias and influence in the data analysis as an in-service teacher. I was purposeful in my participant selection to fully maintain anonymity and the participants’ ability to speak about their experiences freely. Additionally, purposeful steps were taken during interviewing and analysis to increase the trustworthiness of the data and results.

**Participant Recruitment**

As a practicing high school science teacher, I considered recruiting participants from the high school where I work. Initial interview protocols were piloted with some of my work colleagues. During the pilot interviews, I realized my colleagues and I often made assumptions about shared understanding or perceptions of assessment and grading because of our familiarity working with one another over several years. Additionally, there was some hesitancy to acknowledge grading practices that were outside of or subverted district expectations.
I adamantly wanted this research to give a larger voice to other teachers without being filtered through my own views. Ultimately, I decided the best way to objectively answer the research questions was by recruiting participants I did not know personally who were educators from outside my own school district. These participants would not assume a shared understanding of the school or district culture or district expectations with me. Additionally, I had no pre-conceptions about their teaching practices or district expectations, which minimized potential research bias.

**Interviews and Member Checking**

Throughout the course of interviewing participants, I engaged in frequent member checking. This occurred informally during interviews as I often rephrased what I thought participants said and asked for confirmation or clarification. Additionally, during the second set of interviews, after preliminary analysis of the first interviews, I questioned participants about statements or themes that emerged from the first interview. This provided a second layer of member checking.

In addition to formal and informal member checking, both rounds of interviews used semi-structured interview protocols to ensure topics were covered consistently across participants. Follow-up questions were asked as needed but, ultimately, I stayed close to the protocol questions. When I finished interviewing participants, they often had questions about the purpose of the research, my personal plans, and how my year of pandemic teaching was going. However, none of this post-interview discussion was included in the data or used in any subsequent analysis.
Fieldnoting and Memoing

As stated earlier in this chapter, memoing was used frequently throughout the research process and included both a physical research notebook to capture written notes by hand and using the memoing tool within the ATLAS.ti coding software. When I was still planning out the study, I journaled frequently about possibilities for data collection, in addition to reflecting on my role as a researcher and a practicing teacher. During data collection and interviews, I kept a research notebook and wrote fieldnotes both during interviews and afterwards. While listening to transcripts, I wrote about my initial impressions of the data, potential coding categories and areas to explore in more detail. As research progressed, I often memoed about the research process itself, in addition to findings, with the goal of creating a research process audit trail.

Feedback from Others

At multiple stages in this research, I discussed and received feedback from members of my doctoral committee. We discussed multiple aspects of both data collection and analysis including initial coding ideas and preliminary findings. This peer review process served as a prominent additional step to minimize researcher bias and strengthen the validity of the findings and conclusions.

Summary

In this chapter the research questions were introduced and a rationale was provided for selecting a phenomenological perspective and qualitative, interview methods for this study. Participant selection criteria and recruitment were also discussed. Data collection, coding, and analysis steps were then described. Lastly, the researcher’s role in data collection and sense-making process were outlined as well as the steps taken throughout the research process to minimize researcher bias were detailed. The next chapter describes the initial findings from the
research focusing on the policy context related to grading in Maine, the school district context and grading policies, and perceptions of the science teacher participants related to their grading practices. A second findings chapter explores themes of common teacher practices in handling borderline grades across the different participants and school contexts.
CHAPTER 4

REPORTING OF FINDINGS: GRADING CONTEXTS

The goal of this study was to explore how secondary physical science teachers make decisions about borderline grades and what factors influence their decision making. In this chapter, the study schools and participants will be described in detail. Participants were selected based on specific attributes of their teaching role and experience, rather than the school or district they taught in. One goal of this research was to understand the school and district context in which participants made decisions about borderline grades.

The broader contexts of secondary education in Maine will be briefly described in this chapter with respect to the state’s education policy related to student assessment, secondary school choice in Maine, and how school districts are structured in the state. After contextualizing the larger state context, the study schools and science teacher participants will be described. Teachers’ perspectives on their schools’ grading policies and expectations will also be discussed and used to characterize different types of grading systems that exist at the study schools.

State Context

In May 2012, the state legislature passed L.D.1422, An Act to Prepare Maine People for the Future Economy. This new law required all Maine schools to assess students on demonstrated proficiency of the state’s learning standards, the Maine Learning Results (MLRs), in eight different content areas. By 2018, it was expected that all schools would be in compliance and award proficiency-based high school diplomas. During this period, some districts completely transitioned to proficiency-based grading and diplomas. Other districts maintained traditional grading or used a hybrid model, in which numerical and letter grades were still kept while proficiency was demonstrated through various assessments (Stump et al., 2016). Schools were
left to determine their own proficiency standards and how to assess them (Stump et al., 2016). For high school science instruction, many school districts adopted the Next Generation Science Standards (NGSS) Science and Engineering practices or a modification of the practices. In 2018, new legislation passed that would allow individual school districts to choose their high school graduation requirements, rather than enforce proficiency-based diplomas statewide. Some school districts kept the proficiency-based model they had developed, while others opted back into the traditional grading scale they were or had been using to grant diplomas or even some mix of the two approaches.

At the time of L.D.1422 and currently, there is no state mandated grade or proficiency tracking and reporting software. School districts are able to choose their own system to track students’ course and assignment grades. The two most commonly used systems are Infinite Campus (IC) and PowerSchool. The grading software teachers use to record grades is not of particular interest to this study. However, each software platform has embedded programming that does or does not round grades, which may be relevant to this study and teachers’ grading practices, particularly in regard to borderline marks. Teachers spoke frequently about this in the interviews and recognized their ability to override the software default.

**Sample Schools**

In this section, each school in the study will be described briefly. Their grading structure or policy and changes from L.D.1422 and proficiency-based diplomas will be described as well. The twelve participants in this study were from six different high schools (grades 9-12) in six different school districts. Given that Maine is a rural state, it is very typical to only have one high school per school district, unless those districts are quite populated (e.g., Portland) or consolidated.
Table 4.1 School Sample

<table>
<thead>
<tr>
<th>School Pseudonym</th>
<th>Approximate Enrollment</th>
<th>Grading Scale</th>
<th>Total Number of Science Teachers</th>
<th>Number of Study Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marsh Valley HS</td>
<td>700</td>
<td>Numeric, 7 point</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>Woodland HS</td>
<td>1000</td>
<td>Numeric, 7 point</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>Lakeview HS</td>
<td>350</td>
<td>Numeric, 10 point</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Hilltop Academy</td>
<td>500</td>
<td>Numeric, 7 point</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>Riverside Academy</td>
<td>500</td>
<td>Numeric, 7 point</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>Seacoast HS</td>
<td>550</td>
<td>Proficiency</td>
<td>7</td>
<td>2</td>
</tr>
</tbody>
</table>

Marsh Valley High School serves approximately 700 students in an affluent coastal city in southern Maine. At the time of this study, the school used a traditional seven-point grading structure, as did three other schools, as described in the table below. In this traditional scale, students earn course credit through grades and demonstrate proficiency on large assessments throughout their science classes to meet graduation requirements. Grades and proficiency were tracked by teachers and parents using PowerSchool. Three of the participating teachers in this study worked in this school.

Table 4.2 The 7-Point grading scale used by Marsh Valley HS, Woodland HS, Hilltop Academy, and Riverside Academy

<table>
<thead>
<tr>
<th>Point Range</th>
<th>Letter Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>93-100</td>
<td>A</td>
</tr>
<tr>
<td>85-92</td>
<td>B</td>
</tr>
<tr>
<td>77-84</td>
<td>C</td>
</tr>
<tr>
<td>70-76</td>
<td>D</td>
</tr>
<tr>
<td>0-69</td>
<td>F</td>
</tr>
</tbody>
</table>

Parental pressure for students to do well is intense in this community. One participant described a jarring experience upon switching to this school district compared to her former school:

As a, as being new in the district, and you know, having to go through all the, you know, the gauntlet of being approved over those first three years by administration, but also the
community. You know, you got to like fit in. It - I definitely wasn't going to keep doing it that same way. It was not worth it.

According to teacher interviews, parents at Marsh Valley expect students to get As and Bs and often advocate for their children in a very active manner, through emails to teachers, social media and meeting with teachers in person.

Woodland High School is another school district in southern Maine in the coastal region. Three of the participants in this study work here. Woodland High School serves about 1,000 students. In an effort to obtain accreditation from the New England Association of Schools and Colleges (NEASC), the high school shifted to a standardized method of grading for all subjects. In this system, summative assessments counted for 75% of a student’s grade. Any summative assessment with a grade under 80% was allowed to be retaken or redone for up to an 80%, but not higher. Teacher discretion could be used both in determining what assignments count as summative assessments and how these assessments are redone. Formative assessments counted for 15% of student grades. The remaining 10% of the course grade was based on students’ 21st Century skills (Trilling & Fadel, 2009), which is assessed using a school-wide rubric. Science teachers were required to assess these skills every few weeks. There was no minimum or maximum number of summative or formative assessments a teacher could give and report during a quarter or when during the courses these need to occur.

Woodland High School also used a traditional seven-point grading system where students earn both a numerical and letter grade. Infinite Campus was used for grade tracking and reporting. The teachers at this school did not discuss if and how any type of proficiency assessment or score was used to meet graduation requirements.
Students who fail core courses with grades between a 60 and 70 were given “prescriptions” to do over the summer to obtain a passing grade and course credit. They worked on specific course content they needed to complete or obtained passing marks during the summer or as part of summer school.

Lakeview High School is a small school and district in central Maine. The high school serves roughly 350 students during an academic year. Lakeview High School uses a ten-point grading alphanumeric scale as described below.

<table>
<thead>
<tr>
<th>Point Range</th>
<th>Letter Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>90-100</td>
<td>A</td>
</tr>
<tr>
<td>80-89</td>
<td>B</td>
</tr>
<tr>
<td>70-79</td>
<td>C</td>
</tr>
<tr>
<td>65-70</td>
<td>D</td>
</tr>
<tr>
<td>&lt;65</td>
<td>No Credit Earned</td>
</tr>
</tbody>
</table>

Like Woodland High School, Lakeview High School assigned numeric scores to students’ habits of work. However, these scores did not factor into a student’s course grade. Yet these scores could affect a student’s eligibility to participate in extra-curricular school activities, like sports or the school play.

There is another important aspect of schooling in Maine to discuss prior to describing the next two schools. Maine is a rural state and many smaller towns do not have a high school specific to that town but support the cost of a regional high school with other neighboring communities. Some school districts only support the tuition cost of students who attend the specified district high school, while other districts have a policy that allows parents a choice about which high school their children will attend, and town taxes often go towards the cost of school attendance.
The next two schools that will be described serve students from both the town/city in which they are located as well as students from outlying districts. One of these schools, Hilltop Academy, is fully private: students must apply, be accepted, and pay tuition. Unlike Hilltop Academy, Riverside Academy, is “semi-private” secondary school that is supported by local public taxes and private tuition. Riverside Academy serves as the regional school for its geographic district. Like other schools in the state, students living in this district attend without paying tuition because this is covered by municipal taxes. Students in outlying districts (without high schools) with school choice may also attend and municipal taxes would pay their tuition. Only outlying students need to apply for admission or be accepted in order to attend.

Both of these schools serve a moderate number of students, with approximately 500 students attending Hilltop Academy during an academic year and 700 students attending Riverside Academy.

Both Riverside Academy and Hilltop Academy use the common seven-point grading system as seen in other schools. The three teachers interviewed from these two schools felt they had freedom in deciding how they set up the grading structure in their classroom and with regards to the grades they assigned. There were few administrative guidelines for grades. For example, course grade percentages were common throughout the school. Administration required set percentages for how much quarter grades and midterm/final grades would count towards the semester grade for the course: 40% for each of the two quarters and 20% for the midterm (for first semester) or 20% for the final (for second semester).

The last school in this study is Seacoast High School. Approximately 550 students attend the high school during a given school year. Unlike the other schools, this school shifted completely to a proficiency-based diploma when the 2018 mandate was in effect. The two
science teachers in this school graded students’ proficiency on the eight NGSS Science and Engineering Practices. Regardless of the course content (physics, chemistry, etc.), all students were assessed and graded only on these practices. Students were given a one, two, three, or four on graded work. A score of one corresponds to not meeting the standard. A score of two partially meets the standards. A score of three meets the standard and a four exceeds the standard.

Teachers designed tests, quizzes, or other graded work to assess one or more of the eight standards throughout the course. Each of the eight standards were averaged over the course of the marking period and then all eight averages are averaged together. The grade management software, PowerSchool, then recorded and averaged student proficiency scores to give a numerical score between one and four, recorded out to the hundredth decimal place. The software rounded up at particular quarter marks per district requirements and report card formatting. For example, a 3.25 course average would be rounded up to a 3.5 on a report card. Additionally, the numerical threshold for meeting and exceeding standards are different for report cards and transcripts then for general classroom grading. In the district, a 2.5 or 3 constitutes a meets the standard and a 3.5 or 4 exceeds the standard. In order to complete graduation requirements for science, students must pass (earn a 2 in) a certain number of science classes and demonstrate proficiency in the eight science practices at some point in their high school career.

**Science Teacher Participants**

Twelve teachers from six different high schools in Maine participated in this study. They agreed to be interviewed via Zoom twice during the 2020-2021 school year. During the first interview, background information about each teacher was collected, along with their philosophical perspectives on grading and assessment and their experiences teaching during a
worldwide pandemic. For each type of school grading structure, the teachers as individuals will be described in this section, as well as their perspective on the type of grading structure their school used. Pseudonyms are used for the participants and their schools through this dissertation. The first letter of each participant’s pseudonym corresponds to the type of school they teach which will be described fully in the subsequent section. Pseudonyms that start with the letter “P” are teachers from schools using a proficiency-based grading policy. Pseudonyms that begin with “M” or “F” are from schools that used numerical grading with “many” or “fewer” guidelines for teachers, respectively.
Table 4.4 Science Teacher Sample

<table>
<thead>
<tr>
<th>Teachers Pseudonyms</th>
<th>Schools’ Guidelines on Grading</th>
<th>School</th>
<th>Years of Teaching Experience</th>
<th>Physical Science Courses Taught During 2020-2021 School Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Piper</td>
<td>Proficiency</td>
<td>Seacoast High School</td>
<td>7</td>
<td>Physics, Chemistry</td>
</tr>
<tr>
<td>Paul</td>
<td>Proficiency</td>
<td>Seacoast High School</td>
<td>24</td>
<td>Honors Physics, AP Physics</td>
</tr>
<tr>
<td>Mary</td>
<td>Many District Guidelines</td>
<td>Woodland High School</td>
<td>20</td>
<td>AP Chemistry, College Preparatory Chemistry</td>
</tr>
<tr>
<td>Martin</td>
<td>Many District Guidelines</td>
<td>Woodland High School</td>
<td>37</td>
<td>College Preparatory Chemistry</td>
</tr>
<tr>
<td>Melissa</td>
<td>Many District Guidelines</td>
<td>Woodland High School</td>
<td>16</td>
<td>College Preparatory Physics, Honors Physics</td>
</tr>
<tr>
<td>Flynn</td>
<td>Fewer District Guidelines</td>
<td>Marsh Valley High School</td>
<td>16</td>
<td>Freshman Physics, Chemistry, AP Physics</td>
</tr>
<tr>
<td>Fiona</td>
<td>Fewer District Guidelines</td>
<td>Marsh Valley High School</td>
<td>19</td>
<td>Freshman Physics</td>
</tr>
<tr>
<td>Felicity</td>
<td>Fewer District Guidelines</td>
<td>Hilltop Academy</td>
<td>9</td>
<td>Chemistry, Honors Chemistry</td>
</tr>
<tr>
<td>Fallon</td>
<td>Fewer District Guidelines</td>
<td>Riverside Academy</td>
<td>15</td>
<td>Earth Science, College Prep Physics, AP Physics</td>
</tr>
<tr>
<td>Fran</td>
<td>Fewer District Guidelines</td>
<td>Hilltop Academy</td>
<td>20+</td>
<td>Chemistry, AP Chemistry</td>
</tr>
<tr>
<td>Fern</td>
<td>Fewer District Guidelines</td>
<td>Marsh Valley High School</td>
<td>18</td>
<td>Freshman Physics, Honors Chemistry, AP Chemistry</td>
</tr>
<tr>
<td>Fred</td>
<td>Fewer District Guidelines</td>
<td>Lakeview High School</td>
<td>43</td>
<td>College Preparatory Physics, Honors Physics</td>
</tr>
</tbody>
</table>

Three Approaches to Grading Policy and Practice

Each school’s approach to grading policy can be placed into one of three different categories. These categories are based only teachers’ descriptions and perceptions of grading expectations at their schools and their self-reported ability to make decisions about grading.

First, these schools differed based on how student grades were communicated – either on a 0-100 scale or a proficiency-based 1-4 scale. This first difference separates Seacoast High School from the rest of the schools. Within the schools that use a numerical 0-100 scale, another important difference emerged – the relative amount of official guidelines for grading and grade reporting set forth by the school district. Woodland High School has the most prescriptive rules for how
teachers grade in the classroom – both with predetermined grade percentage allocations and specific classroom guidelines. Despite having these large, overarching guidelines in place, some smaller aspects of grading in that school were left to teacher discretion (e.g. – how to administer a summative assessment retake).

The remaining schools can all be categorized as having fewer district guidelines for grading in the school and classroom. Teachers were primarily left to design their grading systems and assign grades as they see fit. Despite differing in the nature of the school (public versus private), these teachers experience a similar environment within their grading.

Figure 4.1 Approaches to Grading for the School Sample

This categorization is based solely on the official guidelines placed on grading. It is worth acknowledging that schools within the “few district guidelines” category did have different school cultures and most likely unofficial school norms about grading. However, due to the nature of this study, this is an area of both limitation and potential future research. Teachers with pseudonyms starting with “P” worked at proficiency-based schools. Teachers with pseudonyms starting with “M” worked at schools that used a numerical grading scale and had
many guidelines about grading in place. Teachers with pseudonyms starting with “F” worked at schools that also utilized a numerical grading scale, but had relatively few grading guidelines.

**Proficiency-Based Schools**

The single proficiency-based school in this study, Seacoast High School, had the smallest number of participants with only two participating teachers, Piper and Paul. Piper was recruited first and helped recruit Paul. Piper and Paul had worked closely together teaching a freshman physical science course for several years. They both spoke about co-designing assessments and practicing using rubrics to grade cooperatively and consistently with one another. During the 2020-2021 school year, however, their teaching assignments did not overlap.

Piper had been teaching at this school for the duration of her teaching career. She had only used a proficiency scale to grade students and had never used the 0-100 scale in her teaching experience. Grades, to her, were a method of reporting student achievement to students, parents, the school, and the state. Piper seemed fairly comfortable with the grading system in place, as she was unfamiliar with any other possibility. Her largest concern surrounding grading and assessment was developing rubrics with clear and articulate expectations and parameters.

Paul had been teaching at Seacoast High School for over two decades, with several years of teaching experience prior to that. Unlike Piper, Paul was at the school when the grading structure switched to proficiency-based grading. He had adopted a proficiency-based mindset – where he focused on assessing students based on how well they can perform the NGSS science and engineering practices. He was able to clearly articulate some of his own challenges surrounding the switch and negative feelings about how the policy has been implemented. He has found it challenging to differentiate between “meeting the standard” and “exceeding the standard.” In the past, he felt students needed to exceed the standard on their own, not with
direction from the rubric. Once expectations were defined in the rubric, the standard could only be met; thereby negating the ability to exceed. After struggling with this for several years, he decided to define the level of “exceeding the standard” and equates it with a letter grade of A; while he equates “meeting the standard” with a letter grade of B. He found the school policy on how to determine a grade concerning, referring to the process as being “bastardized.” Students’ proficiency levels were entered in the grading software as a 1, 2, 3, or 4 and then averaged like a GPA. His concern was philosophical, and he questioned what it meant to average a single standard and then average it across multiple standards.

**Many District Guidelines**

Three of the teachers participating in the study worked in Woodland High School. Woodland High School was categorized as having many classroom and school guidelines for assessment and grading of student work. Teachers reported being regulated in what assignments they could grade and the relative weighting of assignments to calculate the final grade. They also said the high school policy specified the weighted percentages of different assignments that were able to count towards a course grade: 75% of the course grade was based on summative assessments, 15% on formative assessments, and 10% of the grade was based on a school-wide rubric used to assess students’ 21st Century Skills. Teachers were not allowed to take off points for late assignments and students were allowed to retake any summative assessment up to an 80%. From teacher reported information, this is a policy implemented at the high school only; not at the lower grades (K-8). Woodland High School was the only high school in the school district and the only school that implements the common grading for grades 9-12. This policy did not focus on proficiency-based diplomas or grading, and it was unclear from the interviews what content or standards the summative and formative assessments were being used to assess and if
these were school wide, specific to the science department, or content/course specific (e.g., chemistry or physics).

All three of these teachers worked at Woodland High School before this new policy was adopted and they expressed their doubts about the school and district grading policy used at the time of this study. All three viewed the grading policy as lenient and provided students with “an unbelievable” (Martin) opportunity. All three expressed surprise that more students did not take advantage of the opportunity to redo summative assessments to an 80%. Melissa worried that it was “promoting mediocrity” since students did not need to be prepared for summative assessments in the first place.

Mary had been teaching at Woodland High School for approximately 20 years. She valued summative assessments in her classroom as a measure of student learning. Because she taught honors and AP courses, she held student learning to those higher standards and often found students’ summative grade reflective of their understanding. She seemed accepting of the administrative framework on grading summative and formative work. She expressed concern about grading 21st century skills because of the specific point allocation teachers are forced to give. Each skill can be ranked on a proficiency level with a particular point allocation – where the lowest possible grade is a 4. Therefore, a student that did not attend school would have a non-zero grade. Apart from that particular concern, she did not have any objection to the grading system put in place by the administration. She noted that students have so many chances to be successful, she feels comfortable “leaving” the grade in Infinite Campus as is and not changing it.

Martin felt similarly to Mary in leaving grades as is in IC. He frequently stated through both interviews that he felt students had “an unbelievable opportunity” and “one heck of deal.”
He was surprised at how many students do not take advantage of the school policy to earn passing or better grades. He had been at Woodland for almost 40 years and provided his perspective on why the school adopted the change and how it has been implemented. Martin felt irritated that formative assessment did not count for a higher percentage of the grade. He wanted students to be engaged with the formative work and values having students keeping up with daily classwork. Given the current percentages, he felt that students only had to do well on a few tests rather than: “staying in tag, doing what they're supposed to be doing, kind of keeping up.”

Melissa was the most vigorous grader in this study. Unlike any of the other participants, she graded every single piece of student work she assigned; regardless of the class or level she is teaching. She typically spent approximately two hours a day grading student work. She thought students should be given credit for the work that they do: in the form of a grade. Of the three participants at this school, Melissa struggled the most openly with the high school’s requirement of weighting of grade categories. In particular, she felt the formative work that students completed counted for far too little. She felt that a higher weight for formative assessments in the overall grade calculation would place more value on and importance of practice work. She also worried that having too few summative assessments count for so much of the course grade could be detrimental for students. Melissa wanted students to have multiple opportunities to show their understanding. She enthusiastically and positively acknowledged the initiative in the science department to increase the weighting of formative grades to count for more of the overall course grade.

Woodland High School enacted a district policy to grade more uniformly at the high school level. However, variation in grading practices among participants clearly existed. The participants at this school valued different aspects of teaching and learning in their classrooms.
and did assess and grade differently, despite being within a district grading structure designed to create consistent and similar assessment and grading practices among its high school staff.

**Fewer District Guidelines**

Felicity had been teaching at Hilltop Academy for seven years with nine years of teaching experience overall. Fran also taught at Hilltop Academy and had been working there for over twenty years as a chemistry teacher and several prior years other subjects. Felicity taught closely with Fran. They both taught sections of the same regular-level chemistry classes. Felicity taught honors chemistry students that fed into the AP class taught by Fran. They both worked closely on creation and grading of assignments, labs, tests, midterms and finals. With a few exceptions early in their careers, both teachers felt that they had a lot of freedom in their grading and autonomy in their classrooms.

Fallon had been teaching at Riverside Academy and has for approximately fifteen years. She experienced relative freedom in her grading, in deciding what and how to grade. During the Covid-19 pandemic, the administration was more direct in their guidance on grades and credit. Other participants noted this as well. However, Fallon felt she was allowed to assess and grade autonomously, and thus perceived fewer rules about grading in her district’s policy or approach. She mentioned feeling a high degree of parental pressure surrounding student attainment of high grades. Although she recognized the presence of this pressure, she did not mention frequent instances of challenges from parents or administration concerning grades. In her second interview, she spoke in detail about a potential conflict that could occur, indicating she had a good understanding of both administrative and community expectations.

Flynn, Fiona, and Fern all worked at Marsh Valley High School. They all taught at least one section of a freshman physical science class along with other classes. All of these teachers
felt they had significant autonomy in deciding what student work to grade and how to grade it. As teachers of the same freshman physical science course, they made a point to cover curriculum and to grade consistently, so all students had a similar experience. Fiona and Fern were closely involved with each other in designing curriculum and assessments and making course decisions. Both adamantly stated that administration was not making them do this; rather it was a decision made in the best interest of students: “…we have to collaborate and be in unison. We don't have to. Nobody actually told us we had to, but we feel like that's fair.”

Fiona and Fern explicitly categorized their school as having significant parental pressure surrounding the attainment of “good” grades. Flynn alluded to this pressure indirectly. When asked about raising or not raising grades, he made several comments about certain things “not being worth it.” Teachers at this school had several common assessments students had to take as part of L.D.1422. Students needed to demonstrate proficiency in the NGSS science and engineering practices. Demonstrating these proficiencies was a graduation requirement; however, the teachers were largely left on their own to decide what and what to grade for individual student work and how to determine course grades.

Fred was the most experienced teacher in this study with over 40 years of teaching experience, most, if not all, at Lakeview High School. He had been at the school long enough that he had taught children of former students. He had an established way of “doing things” in his classroom and experienced little to no friction or opposition from his administration, other teachers, or community members. He was incredibly focused on grading students objectively. He worked to establish meaningful relationships with every student and create a classroom environment where students could come to terms with being graded objectively. He believed objective grading to be particularly important for his honors physics students, especially the
He again focused on objectively grading their content knowledge and would not “bump” students’ grades based on effort or personal feelings: “I try to make it as objective as possible…I’m going to try hard to grade you, not on whether I like you or not, but…on the quality of your work.” Fred made it a point to communicate this to students and pushed them to learn the material rather than only trying to get a “good” grade.

**Summary**

This chapter described the state and school contexts in which the teacher participants taught secondary physical science. Three different approaches to grading were taken by the schools in this research: proficiency-based grading, numerical grading with many district guidelines, and numerical grading with fewer district guidelines. Additionally, teachers’ perceptions of grading policies and expectations at their schools was described in depth, looking at both district and parental expectations.

Teachers’ experiences with their district and school grading expectations were characterized into three different types of grading systems: proficiency-based, systems with many grading guidelines, and systems with fewer district guidelines. Most of the study schools had fewer guidelines for school and classroom grading and teachers felt a large deal of autonomy in their decision-making about what and how to grade. However, despite differences in schools’ approaches to grading, teachers were often able to exercise individual choice and modify their classroom and course grading, especially with respect to borderline grades.

In the next chapter, teachers’ grading practices and patterns in determining borderline grades will be examined. Factors that influence teachers’ decision-making about borderline grades will be outlined and as well as the process for deciding to make grade adjustments.
CHAPTER 5
REPORTING OF FINDINGS: BORDERLINE GRADING FACTORS

The previous chapter explored the broader context of the study related to education policy and student grading in Maine secondary schools, demographic information on the schools and teachers in the study, patterns in district approaches related to grading, and teachers’ perceptions about grading expectations and their own grading practices. In this chapter, two key areas of this study will be addressed: the identification of factors that influence teachers’ decision-making surrounding borderline grades and how teachers navigate the decision-making process with these factors in mind.

Grading Practices in Borderline Cases

Several consistent findings emerged from the data regarding the nature of how teachers consider, experience, and make decisions about borderline grades. Across the twelve interviews, all participants responded similarly to particular interview questions, regardless of their subject taught or the school they were teaching in. These trends and supporting evidence will be discussed in the following sections. Evidence will be provided in the form of collective and representative examples rather than examples and quotations from every interview.

Quarter, Course and Semester Grades versus Individual Assignment Grades

The first several questions of the interview protocol were designed to elicit what types of assignments teachers were conscious, contemplative, and considerate of when deciding borderline grades. After being asked about difficult borders for assigning grades, teachers were then asked about what particular pieces of student work produced these conflicts. Across the board, teachers in this study agreed that quarter grades, semester grades, and course grades were the most difficult cases to make decisions when it involved borderline grades. Although teachers
were aware of student grades throughout the course of a marking period, individual class assignments (like tests, quizzes, or homework) were not the source of much deliberation. Furthermore, when teachers were asked about adjusting numerical scores on student work slightly below the border for a mark or grade, they only referenced raising the grades for quarter, semester or final grades, “big” grades as one teacher called it. Felicity stated it clearly: “… for me, there’s only borderline grades when it comes to like quarter grades.”

All of the participants indicated during interviews that these borderline grades warranted attention and potential adjustment because they were of higher perceived importance or had larger ramifications for students. In the teachers’ views, assigning these grades was not a simple tally of points or weighted averages; rather, other things needed to be factored into their decision-making process. Martin described how challenging it was to give a final failing grade that was so close to passing:

You’d be hard pressed to give a kid a 69. We’re talking final grade. I don’t have any problem with regular grading throughout the semester. But if you’re talking the final grade, end of the year… A 69 is kind of tough.

In this particular comment, Martin did not get specific about why assigning this particular final course grade is tough. The determination of the grade seemed to be more than just a mathematical calculation for him. Martin hinted at the emotions associated with giving a student a borderline grade, even if he did not provide details. This suggests that issues beyond straightforward numerical computations are in play when assigning borderline grades that are final course grades. Other factors may be (or seem to be) involved in deciding the final grade.
Considering Consequences for Students

Many participants worried about the ramifications for students related to their decisions about end of term grades. Fran, at Hilltop Academy, worried about the difference between graduating with honors (all As and Bs) or high honors (all As):

…The assignment, I leave it as is. I don’t really play with grades, but look at where they are… at that 92.4. You know, one question on a test moves from B to an A…to get three chords versus two chords or something like that. Yeah, so I don’t do it on individual assignments. I do it at the quarter level.

In considering quarter or semester grades, Fran considered adjusting borderline grades because of what could happen to the student at the end of their high school career. Other participants also worried about potential long-term consequences for students when focusing on end of term grades. Fred did not care about individual pieces of work, but rather final grades: “In general, on a piece of regular work, like I said. If it’s an end of semester [grade], where it’s the difference between passing and fail or something like that, it’s a different story”. In Fred’s view, individual assignments did not warrant special attention and were not modified, which he started earlier in the interview. When he says “like I said” he is referencing how he grades students objectively and their score is their score.

Learning over time

In the view of the teachers, these final course grades had a higher degree of permanence than the individual assignments grades students acquired over the timeline of the class. During their classes, teachers often had structures built into their course so that students could continue to learn, earn points and gain mastery in particular content areas. For most teachers, this included options of assessment (test or quiz) corrections or retakes. The adoption of these policies eased
pressure on the teachers surrounding decisions about borderline grades during regular coursework. Therefore, teachers only needed to make decisions about borderline course grades at the end of a marking period.

For most of the teachers in this study, they had enough district flexibility that this was a practice they chose to adopt. The reasons for these additional opportunities for students to demonstrate their learning were not always clear. However, in some cases, teachers made the case that students needed to continue to work towards proficiency in particular skills or areas. Test corrections and retakes gave them both the opportunity to continue learning and to demonstrate proficiency. Fern shared this comment:

…rarely would I just change an exam grade, because I allow students to do work...to demonstrate that they're, you know, still continuing to work on proficiency in those skills…And if they can explain what they got wrong, and how to do it correctly, then, then I will raise it up, but, I give them a way to do that. So it's an option and opportunity…

In the school with many district grading guidelines, Woodland High, the opportunity to retake summative assessments was a specified district policy. It is unclear what practices the teachers at this school implemented before the district mandate and if they had previously allowed assessment “retakes” or correction options like many of the other teachers in this study. Teachers at Woodland High viewed the current district policy of “retakes” as an incredible opportunity for students and adamantly stated they did not experience any dilemmas when assigning borderline grades to summative assessments over the period of the course. However, they did find end of term grades more challenging for assigning numerical grades. Martin explained,
With tests, I don't, I don't change it. They're just they get what they get. But if it's a semester grade, it depends on what they've done. I look to see if they've done their formative work or not. You know, if they haven't passed anything in or made any effort… I don't feel too bad [in] letting the grade stand. But if they've done all their work… Their attendance is good, their work ethic is good… And I don't feel bad about going from a 69 to a 70.

In some ways, the perceived generosity of the district policy to allow students to retake summative assessments, while sometimes frustrating for the teachers, seemed to alleviate indecisiveness in their decisions about final grades. Additionally, teachers in the school with many grading guidelines (Woodland High School) were also required to refer failing students to summer school opportunities to earn course credit before the following school year. This was not unique to Woodland High School. Marsh Valley High School also had this opportunity, but it was less of a formal requirement or expectation.

In general, teachers experienced the most challenge in deciding on borderline grades when they occurred at the end of a marking period. This is not to say that teachers did not bring up instances of challenging or surprising grading experiences in other circumstances. For example, teachers in this study mentioned being surprised in mainly two situations: when the whole class did poorly on an assignment or assessment, or when an individual student did more poorly than expected (compared to their past performance). However, these situations did not strike teachers as challenging borderline grades to assign. In the instances described, the grades were not necessarily borderline; they were attention-catching. In both scenarios, the teacher revisited the assessment to determine the cause of the anomalous results. The latter scenario is described in more detail in subsequent sections.
“Bumping Up” but never “Bumping Down”

One question in the interview protocol asked about numerical grades that were just above the border for a mark, letter grade, or proficiency level and asked if these borderline grades were ever adjusted. After several interviews, the theme of “bumping up” but never “bumping down” a grade became clear.

All the teachers were surprised by the question and asked for clarification; specifically, asking if the question was about lowering grades or bumping students down. When my response confirmed that lowering the score was a possible adjustment a teacher might make, all the teachers agreed that they did not engage in this practice. Additionally, their responses were curt and adamant about not engaging in the process of lowering grades even when pressed with follow up questions. Felicity shared a typical response: “I never adjust it. No, I honestly, when I go down my gradebook, I'm only looking for scores that are on that border. Below, I should say, below that border.”

Teachers did not bump down grades for two primary reasons: (1) an ethical concern about purposefully lowering a grade below the average calculated by the grading software and (2) the perceived logistical difficulty of modifying the grade. However, the ethics and morality involved in were not well articulated in teachers’ responses. Rather, these moral objections were evident in how teachers responded to the interview prompt, both with their verbal and non-verbal communication cues. As stated earlier, teachers responded quickly and curtly when asked about if and when they engaged in lowering student grades that were above the border for a mark. In responding to the interview question, many teachers adamantly answered that not only did they never lower grades, they never would. When asked about lowering grades, there was the underlying expectation that they did (or could) engage in that behavior. When teachers used
negative statements in response to the questions, they were indicating a violation in the expectation couched in the interview question – again, that they could or did engage in this practice of lowering grades. Additionally, many teachers responded with surprise that the question was even asked. Their surprise and short response also indicated a violation of expectations and social norms surrounding teaching behavior. Other interview questions that were within “typical” teaching norms were answered with a more relaxed tone and in greater detail. In contrast, this interview question garnered short responses followed by a pause or silence. Silence can be used by speakers to indicate a completion of talking, signaling turn-taking. Perhaps participants were indicating their readiness to move on to subsequent questions.

Additionally, multiple teachers responded that to change a grade, one would need to go through the gradebook and modify individual assignments to bring the overall grade average down. Mary summarized this experience:

I typically do not adjust grades. So I would say that I let that score stand. I mean, I would find it really hard to adjust it down, because then I'd have to change other grades.

Because we do our grades in IC and the parents can see it, it's very, very transparent.

As stated before, “bumping down” was not considered a “normal” behavior that teachers engage in. The uniformity of teacher reactions about this practice indicates this, but so does the perceived “transparency” of grades that Mary addresses. Both commonly used grading programs, Infinite Campus and PowerSchool, allow parents and students to see their course and classroom grades throughout a term period. While this feature allows communication and “transparency” in grading, it also allows a certain level of oversight into teachers’ grading and marking of assignments. This oversight may constrain teachers’ grading practices, particularly with respect to grades once they are entered into the software.
This behavior provides an insight into teacher grading and societal and community expectations, and the element of public scrutiny or oversight over grading. Gradebooks are expected to be accurate to a point; parents can see a student’s scores for a class over the course of a semester or year. Teachers did not find it acceptable to bump down, but do find it acceptable at times to bump up and end of term grade. As Felicity stated earlier, she only looked for grades to bump up during end of term grading. And, like the other teachers in this study, she adamantly does not bump students down. The upwards modification of the grades, despite having perceived “transparency” and some level of accuracy, is socially acceptable by parents, teachers, and students. Teachers, students, and parents treat the grade presented by electronic grading systems as the minimum grade a student has earned. This cultural perception of electronic grades representing the minimum grade could be further supported by the insistence of participants on not engaging in the practice of lowering grades and offering little to no justification as to why they did not engage in this behavior, other than the difficulty of changing grades already entered and communicated through the grading platform.

This practice of bumping up grades, but not bumping them down, speaks to a larger issue with grading that is not addressed in previous research. Assessment and grading exist within an environment that is specific to the school, district, and broader culture. Norms about grading could exist within schools, districts, the state of Maine, population subgroups, or at the national level. All the teachers in this study normalized “bumping up” and engaged in this practice, but adamantly did not “bump” student grades down. Yet they did not specifically point to or locate the norms that might explain this behavior. This implies that teachers’ behaviors related to bumping up but not bumping down tap into some broader cultural norms and expectations, and that the behaviors may not be a school or district level phenomenon; rather, it may be a broader
trend in education practice. It is very likely this is a cultural facet of teaching and grading in American culture.

**The Importance of Student Effort in Deciding Borderline Grades**

Previous research as described in the literature review chapter, indicates that teachers often consider a student’s effort (and other non-achievement factors) in making decisions about borderline grades. All of the teachers in the current study indicated that they would bump student grades up for certain borderline grades, such as end of term grades, based on the amount of effort a student put forth. They stated this explicitly in most instances. When describing instances when they would raise student grades, they described the various ways in which a student could exhibit effort to learn the material: coming in for extra help, taking advantage of test corrections or retakes, attending school or “doing what they’re supposed to be doing.”

The different categories of effort and behavior described by Randall and Engelhard (2010) were indistinguishable for the high school teachers in this study. Given the unclear definition of what constitutes “behavior” there is a challenge in delineating what clearly counts as effort and what counts as behavior. None of the participants discussed any student behaviors not related to trying to learn course material, such as disrupting class, speaking out of turn, or being obnoxious. The student behaviors these teachers are looked for, as they decided on final course grades, were behaviors that indicate the student is expending effort on the assigned tasks or assessments and trying to do well in the course.

**Borders that Warrant Attention**

Teachers were asked about their experiences and decisions about borderline grades in a variety of settings and borders. Teachers might have referenced particular borders they paid attention to, but during the course of the interview they were continually pressed to consider and
comment on other borders as well. For example, if a participant discussed challenges surrounding the pass/fail border, they were asked if any of those challenges occurred at different borders, i.e., A/B. Two borders emerged as prominent among the teachers: the pass/fail and A/B borders. Although all the teachers treated the pass/fail border similarly and made similar decisions, the A/B border produced much more variation in how teachers experienced and made decisions about those grades.

**The Pass/Fail Border.** When teachers were asked about what particular borders of grades they were most concerned about, there was a universal concern for the pass/fail border. Although the cutoff for failing was different in different schools, teachers unanimously expressed concern about the student and the consequences of failing a course. The teachers in this study questioned whether it would be in the best interest of the student to have to retake the course, attend summer school, or be deficient in credits. Fern shared, “…I wanted so bad for this kid to pass, because I didn't think it was going to be a healthy experience for this child to have to spend summer doing also distance work to recover this credit.”

Additionally, there were other factors besides worrying about students’ well-being that factored into teachers’ decision-making. Course grades that were one to two points below the border were especially challenging to deal with for various reasons. Flynn commented that he did not believe his grading practices and methods of numerical scoring were accurate enough to choose not to round. “If it's that close, I don't feel like my grading is perfect enough that I can be that precise.”

Martin alluded to district pressures and expectations about failing students as being a factor in deciding to round up: “But for a semester grade it's tough…or an end of the year. We try to discourage 69s when somebody's that close.” At this particular school, the cut-off for
passing was a 70. Martin hinted at some level of shared expectation in his use of the plural first-person language: “we try to discourage.”

The further a numerical score got from the pass/fail border, the more often and easily teachers let the grade stand. Flynn, who spoke about precision in his grading, said that he felt comfortable letting grades stand that were five or more points below passing. In that situation, he felt confident in his grading accuracy and that the student did not achieve a passing grade within error. Additionally, like many other teachers in the study, he had opportunities for students to retake assessments to improve their learning and overall course grade. When students were further away from the pass/fail border, this teacher felt absolved of his responsibility to adjust a grade and passed the onus to the student.

**The A/B Border.** Although all teachers reacted similarly in their decision making about the pass/fail border, they differed in how they made decisions about the A/B border. Interestingly, teachers were split on if and when they would bump up students right below the “A” border, regardless of what particular numerical cutoff that was at their particular school. Several teachers refused to bump students up right below this border for various reasons ranging from simply “they get what they get” to wanting to reward students for effort or wanting to teach students life skills.

Two teachers at Woodland High School, which had significant grading guidelines and policies, did not raise student grades up to the A level in the A/B borderline cases. They felt that students already had many opportunities to improve their grade throughout the course, given that summative assessments with grades up to an 80% could be retaken. One teacher mentioned that the reliance on summative assessments as the largest portion of a student grade was a good indication of their overall understanding.
The third teacher at Woodland High School did engage in the practice of raising high B level marks to the A minus range based on student effort. This teacher felt that students needed to be rewarded for engaging in the learning process and formative assessments. She was particularly upset that formative assessments at the school counted for so little and there was no penalty for late work. She engaged in the process of bumping students up based on their effort. She viewed student effort as the timely completion of work and completion of all the work students were assigned.

Several other teachers engaged in this type of “bumping” for the A/B borderline cases. Teachers would look over student performance and make a value judgment based on how they perceived student effort, but also other student characteristics. Students who were invested in learning, turned things in on time, and took advantage of opportunities to improve their grades were often bumped to the “A” mark if they were within a point or two. Fallon explained,

I look at them as a whole student, you know, not even just their ability, their content ability, but I also look at, like, you know, their effort, their timeline, timeliness and with their assignments, even their ability to ask for help when they need it, because I think that that is an important skill…that definitely goes in my final decision about like, their big grade.

In teachers’ views, student effort was entangled with the ability of the students to learn the required course material. The process of learning requires effort. Although effort increases the amount of learning and academic achievement a student can obtain; during the case of the borderline A/B grade, some students are not able to reach the “A” range of work based on their academic achievement alone. Some teachers, like those mentioned above, raise scores for students exhibiting the characteristics they value, like effort. Fran commented,
But if they don't put the effort in…that's on the kid. Because some kids don't have the ability. And that's my job to help them learn. And if I feel that they're that they're receptive to that fine. And their grade will show that. But if they don't even want the help and they don't want the effort, then I have to give them what they earned.

In the example above, this participant exhibited many of the patterns discussed throughout this section. When students exhibit effort and productive learning behaviors, teachers said they would often round up in the case of borderline quarter or course grades. Effort is, by and far, the most documented non-achievement factor in teacher decision making surrounding borderline grades in the literature. Additionally, despite a student exhibiting low effort and low academic achievement, this teacher indicated she would let the grade stand rather than bump it down.

In the proficiency-based school in this study, Seacoast High School, the A/B border was generally deemed equivalent to a proficiency level of “exceeds the standard” or a level 4 versus “meets the standard” or a 3. Paul explained: “…basically a four has turned into like the A.” Both teachers at this school would tend to round up grades when students were on borders, especially for the A/B border. Their methods of rounding up were particular to the grading software they used and how the software and school district policy determined proficiency levels on report cards. On a report card, a 3.5 average for the NGSS standards was equivalent to “exceeds the standard”, rather than requiring a score of 4 used during classroom grading. Additionally, the software had rounding defaults built into it that teachers were cognizant of. Paul described this software constraint:
Yeah, let's see, it's like a 3.2. It's hard to explain (laughing). So if you get a 3.25 and above, then on the transcript, it comes out as a 3.5. And so if they're on the cusp of one of those, then I would bump them up to the next one.

Piper focused more on growth rather than maneuvering around the rounding programmed into the grading software. If a student was able to demonstrate improved performance, this teacher would raise the final grade:

I might have a student who has like, you know, a 3 and then a 4 and a 3.5, and a 4, you know, and I might not take the average, and I might…bump it up to the 3.5 or something like that.

Some of the teachers in this study did not round or bump up student grades that were on the cusp or borderline of district marks for a grade of A. These teachers all were at schools that utilized numerical grading. However, there was no clear pattern of not rounding across the schools with either more grading guidelines or fewer guidelines. Two of the three teachers at Woodland High School, where there were many grading policies, did not bump up student grades. Two of the six teachers in schools with fewer district guidelines did not bump up student grades either. Most of the reasoning provided for not rounding was similar between the teachers; they often stated the grade was what it was. When asked about bumping up on borders other than pass/fail, Flynn stated: “Yeah, I don't really. I don't really do much with it…it is what it is.”

Mary voiced a very similar perspective: “I don't generally raise any scores. I put them all in IC and it is what it is.” As stated previously, like the teachers in this study, her attention is mainly on the pass/fail border and on more permanent grades if at all. Additionally, having many specific district policies about grading, Mary felt like the summative assessment retake policy at
the school was incredibly generous and did not feel the need to raise student scores, especially scores in the passing range.

One teacher in this study adamantly did not round or bump up in cases like this. Fred kept A/B students exactly at what the grading level indicated from the software output. However, unlike the other teachers in this study, Fred focused more on the future life experiences for the student rather than what the student did or did not take advantage of during the class. He considered the adverse effects of rounding up grades for the future academic career of his students,

I think that's a service. Because that's going to happen to almost everybody at some point. And if it happens to you when you're away from home, as a freshman in college, adapting to a million other things, and it's the first time you don't get an A, when you try as hard as you can. It's not good. It's a lot harder. You know, here, they're in a smaller class, where they know their teacher well. And I do a lot of, I spent a ton of time processing with kids. And what I try really hard to get them to think about right from the first day is your goal as a student, and this is, I don't care what level student you are, your goal is to try. Try as hard as you can to learn. To learn as much as you can. Right? And to be able to - see yourself in the mirror every day and say I'm trying as hard as I can. And if you're truly trying as hard as you can, and you get an 88. Great!

Fred was unique from the other participants that did not bump up grades on the B/A border. Other teachers decided not to round up because students failed to take advantage of opportunities to improve their grade (e.g., coming in for extra help or taking advantage of test/summative assessment retakes). Thus, students earned what they earned. Fred also took the stance that students earned what they earned. But, rather than focusing on the past and present
classwork of students in the course to justify the decision not to raise grades, Fred focused on implications for student success in the future.

Many teachers considered potential consequences and future challenges for the student when deciding on how to assign grades for the pass/fail border. These future considerations (e.g., having to retake the course, go to summer school, or not graduate) were considered in tandem with the past and present coursework of the student. In contrast, the teachers that chose to round up or keep the grade as is for the A/B border, primarily considered past and present student work and effort, rather than future considerations. Future implications for students were considered in some cases of raising a grade on the A/B border. For example, one teacher worried about preventing students from making “High Honors.” However, the consideration of future ramifications was far more prominent in how decisions were made about the pass/fail border. Fred was anomalous in that he focused primarily on future consequences when making decisions about the A/B border.

Summary

In summary, the teachers in this study found end of term grades to be the most decision-provoking and challenging. Teachers perceived quarter, semester, and end of year grades to be more consequential for students and often would consider bumping these grades up. However, none of the teachers in this study ever engaged in lowering end of term grades, which possibly is in line with and indicative of broader cultural norms in American schooling. Lastly, teachers paid attention to some borders more than others. The pass/fail border was especially concerning for all the teachers in this study. Again, a large part of this concern focused on teachers’ perceived long term ramifications for the student. Similarly, concerns for the students’ academic future
prompted about half of the participants to perceive the A/B border to be challenging for determining borderline end of terms grades, as well.

**Decision Making Considerations and Processes**

In the previous section, factors that teachers considered when making decisions about borderline grades were discussed. Teachers considered end of term grades to be the “only” borderline grades, focused on specific borders more than others, and often used well documented non-achievement factors, mainly effort, to make decisions about these grades.

The goal of this research was to examine how secondary physical science teachers made decisions about borderline grades. In this section, teacher decision-making will be explored further with two areas of focus that emerged through the thematic analysis: teachers’ concerns with grading as a scientific measurement and perceptions about students’ content knowledge.

In the interviews, teachers’ often voiced concerns with the potential for inaccuracy, bias, and subjectivity in their determination of borderline grades. Teachers’ patterns of decision making will be discussed with respect to each of these three areas. Next, perceptions about students’ content knowledge will be discussed as it relates to decision making on grades. Surprisingly, in many instances of deciding to bump up, teachers focused more on non-achievement student attributes and perceived potential future consequences. When deciding not to round up, despite feeling pressure, teachers often cited students’ lack of content understanding to justify their decisions. The decision to raise or hold end of term grades was indicative of an overarching pattern of behavior which will be called “schema switching.” In deciding borderline grades, teachers often move between different schematic frames in how they approached and perceived grading.
Concerns with Inaccuracy, Bias, and Subjectivity

The secondary science teachers in this study used their content knowledge both to instruct and assess their students. The teachers also used this knowledge to modify grading if needed or make adjustments to their assessments. However, when deciding on borderline or “big” grades, they often resorted to non-achievement factors, mainly effort to determine the final grade. Much of the discussion above has focused on issues that might be common to teachers of any subject.

However, there were comments in the interviews that may indicate some views and behaviors around grading might be related to the fact that these participants were science teachers. These teachers all had undergraduate degrees in a scientific field and thereby, in all likelihood, engaged in disciplinary-related habits during their jobs. It is possible that these scientific habits or perspectives factored into their grading. Specifically, a common concern with inaccuracy, error, and subjectivity in grading came up in many interviews. Teachers’ views about these factors are described in more depth in the following sections.

**Inaccuracy.** When asked about grades that fall just below the border of passing, Flynn justified his decision to bump up by referring to the concept of using scientific reasoning and the practice of measurement error. He recognized grading as a form of measurement with inherent degrees of potential error.

I just rounded up to a 70. If it's that close…I don't feel like my grading is perfect enough that I can be that precise…I'm not sure it's fair to you know, to give somebody a 68 or 69.

When I don't feel like you know, my grading … I'm not sure my grades are that reliable. Flynn acknowledged the lack of perfection and thereby scientific accuracy in his own grading. His grading was not precise. In questioning whether his grading is perfect, Flynn was
acknowledging that a course grade of a 68 or 69 might very well be a 70 within the margin of human error. For most of the schools using the 0-100 numerical grading scale, the cutoff for passing a course was a 70. Later in the interview, Flynn further questioned his own grading and expressed doubt as to how “authentic” his grading was; however, he did not specify what he meant by “authentic”.

Many participants found it difficult to keep grades that fell this close, one to two points below, a failing grade. Fallon said, “I think that the grade of a 69 or 68 is a hard one to assign” and later: “I know that my grades aren't... there's a lot of error in my grading…I think that is a bigger obstacle for me that because I know that my grades aren't a perfect measure.” Both Flynn and Fallon acknowledged that grading has inherent error and that a grade is a subjective and approximate measurement. These teachers focused on their understanding of grades as a measurement of student achievement to help make their decisions. They referenced human error and imprecision, which is at the forefront of their reasoning when making decisions about whether a student fails their class or not.

**Where Inaccuracy Appears.** Teachers did not only generally recognize imprecisions and inaccuracy in grading; they also recognized specifically where it appears. The examples presented above show teachers’ understanding that the numerical score they assign students for individual assignments has some degree on inherent error. Any error existing within these assignments gets compounded over the many grades within a course, which made teachers interpret the final quarter or semester grade with some hesitancy if it fell below the level required for a particular grade mark.
There are nuances to grading that the teachers in this study recognized. Teachers recognized two problematic places in their grading: how they chose to allocate points and, more fundamentally, what they chose to assess.

**Inaccuracy in Point Allocation.** Teachers recognized that individual point allocations for problems on various assessments might not be an accurate measurement of what a student knows. Fern faced this dilemma in her grading often:

Does one skill warrant this amount of like a grade reduction? Because like that would just makes me have to reconsider what my where I've assigned points to the assessment…I've had moments where I might get to one test in particular and say, ‘You know what, I've just taken up too many points for this thing.’ And then I'm going to change my key, my scoring rubric, and I'm going to go revisit the stack that I've already done [graded].

Fern regraded her assignments fairly frequently as she strove for more accuracy in her numerical grading. Fern was not alone in this practice. Several teachers described the experience of regrading or reallocating points on portions of assignments because they felt they were taking off too many points for a given skill. Practicing teachers are often left to determine the point makeup of various assignments as well as the point allocation per problem or section. A comment from Melissa earlier in this chapter detailed her struggle with allocating too many points for a particular skill. Melissa described being surprised when a well performing student got a low grade on a test. Based on her content and pedagogical knowledge, she was able to see a pattern in the incorrect responses of her student on a physics test. In penalizing the student for simply wrong answers, Melissa, like Fern, took off multiple points and instances for the same incorrect skill. Yet Melissa and Fern both modified their point allocation in their numerical grading to better represent what they thought the student should earn.
Inaccuracy and Bias in Deciding what to Assess. Teachers also recognized that there is inaccuracy and potential bias in what they chose to assess. Fallon shared, “Probably there's a lot of error in my grading. And what I assigned [for] things, it's, you know, it can be a little subjective, and what I think is important and not important.” This view was especially apparent for teachers in districts with fewer grading guidelines. These teachers are left to decide the weighted percentages of their grading categories, what gets graded, and how the graded assignments produced numerical scores. However, teachers at Woodland High School, also recognized their ability to decide what gets graded despite having district policies meant to guide and standardize their grading practices.

When Grading is not Representative of Understanding. Teachers also understood that students might have external circumstances that may interfere with their ability to show effort or demonstrate academic performance. Therefore, their scores or grades may not reflect their actual understanding. During most interviews, after teachers explained their grading practices, they added a caveat, explaining that those were their practices unless students had “extenuating circumstances.” Flynn discussed instances where the numerical grade might not represent student understanding: “You know, now, if someone's like a 67, and there's maybe they kind of know their stuff, but there's extenuating circumstances, then yeah, that would bump that up.” These participants recognized that other things happening in the life of a student might not enable that student to effectively show their understanding in their school work.

Subjectivity in Grading. All of the participants often referenced their concern about subjectivity in grading; more often and more frequently than discussing concerns about measurement error or impressions in grading. Paul stated this idea directly: “I would say, like to me, all grading is subjective.” “Subjectivity” appeared for teachers in a variety of different situations: the content
being taught, what is assessed and how it is assessed, and decisions about the student being assessed.

Teachers did not always explain what they meant when they referred to the concept of “subjectivity.” Many of the science teachers felt they experienced less subjectivity in grading science content as opposed to other content areas. Several compared student work they graded to the work of teachers grade in other subjects; speculating that grading in science is less subjective than English, for example. Martin commented, “I could see a challenge as maybe an English teacher or something like that. But in Chem…I don’t have too [much] subjective stuff…it’s problems they need to solve and whatnot. So they’re relatively straightforward to score.”

Many teachers recognized that different types of assignments were harder to score objectively. Essays, which are often used as assessments in English and Language Arts classes, were viewed as being more challenging to grade. Subjectivity could more easily creep into how points were assigned. In addition to essays, teachers also mentioned struggling grading for other types of assignments, like projects or journals. Teachers often used rubrics to assess science projects, essays, or journals and sometimes found it challenging to make sure the rubric clearly communicated expectations for student work. Piper struggled with the ambiguity in how a 3.5 and 4 were delineated on some assignments: “And then sometimes I have trouble maybe distinguishing between a 3.5 or a 4. Because that seems more fuzzy, and I don't have clear communication in the rubric.” Similarly, Mary also found rubrics more challenging to use than numerical grading: “So I feel like when we're grading with a rubric, it's much different. Because if you have a 0, 1, 2, 3, 4… the 3 range is so huge.”

Additionally, teachers recognized what they chose to grade could also be a subjective decision. Fallon was quoted earlier, but her statement bears repeating: “And what I assigned [for]
things…it can be a little subjective, and what I think is important or not important.” Given that teachers can chose what to assess, their choices have inherent bias and therefore, subjectivity. This is true for teachers regardless of the school policy on grading.

Teachers in this study recognized that their feelings about a student could affect how they grade that student. All of the teachers actively tried to avoid having perceptions of students affect their grading. These perceptions could be based on the past performance of the student. Piper shared this observation:

You know, I think that it's easy for a teacher to take a student who normally does well…and even if they didn't do well, in particular test, you can be like, well, you know, I'm pretty sure they meant this instead of what they wrote. So I'll give them a 3…I know I'm not perfect, but I try really hard to follow the rubric instead of how a student normally does.

Teachers also worried about how “liking” a student affected their grading. Paul said he tried to grade objectively, but was honest about the fact that whether or not he liked a student might affect their grade: “If I know a kid super intelligent, and they're just not doing anything, I might tend to think, again, not consciously, it's not conscious…I'm [not out to] get … somebody … [It] just like, I think it just creeps in subconsciously.”

All of the teachers actively tried to avoid the subjectively aspect of grading. Fred worked hard to establish a culture of fair grading based on the student’s work and not on the person.

I’m going to try hard to grade you, no on whether I like you or not, but…on the equality of your work. Knowing that, you know, there are some things…that can be pretty subjective, but I try not to have it be that way.
Fred spoke at length about how he created an atmosphere where students accepted that they would be judged solely on their content understanding. Starting at the beginning of the year, Fred explained,

…What I try really hard to get them to think about, right from the first day, is your goal as a student, and this is – I don't care what level student you are – your goal is to try. Try as hard as you can to learn to learn as much as you can.

Fred later said building relationships with students was key, “If you have a good relationship with if them…you've talked to them and they know you care about how they do. They know that you're trying as hard as you can to be fair with them.” He said he fostered meaningful positive relationships with students, so that he could try to grade as objectively as possible.

Paul dreamed about an alternate grading and assessment setup entirely. To avoid subjectivity, he wanted to avoid grading his students all together by having an outside assessor. He referenced a class he visited at a local University where an instructor taught the students, but an outside person would actually assess the students,

The course worked, where, like, I'm your teacher. I'm trying to teach you how to teach you these skills. And then somebody else assesses if you've met those skills. It's kind of it's similar to an AP test, in a sense, but it's more like it automatically puts the teacher into a, like a coaching position…And it takes the, like, I know, this kid works really hard.

You know, in my mind, I can guarantee you, like, if it comes down to it, and I'm putting the grade on something, the kid that worked hard, if it, you know, is going to get a better grade than the kid who just halfway did something.
In this hypothetical grading system, Paul’s focus would be on instruction, but not on assessment, “I figure out what's the most important things to take away from the class. And then, ideally, somebody else would grade them on how well they made it there.”

The grading behaviors and experiences of these teachers suggested there are larger cultural expectations and norms involved in grading. Grading is not simply assigning an objective numeric score to a piece of work. All the teachers acknowledge subjectivity, but to grade objectively, teachers often feel they need to modify their environment. Paul, ideally, would completely change the way secondary high school teaching and assessment are done in the United States. Teachers would be responsible for only instruction and curriculum development. Assessment would no longer be their responsibility. Fred worked incredibly hard to create an environment where students accept that they will be graded objectively. Paul’s fantasy of a completely different assessment structure and teachers’ efforts to be more objective and fair in grading, implies that they don’t perceive grading as particularly accurate or objective. The fact that Fred needed to work to get his students to accept an objective grading process also implies that students inherently expect some degree of subjectivity in grading. Paul, ideally, would not even grade his students if he could avoid it. Teachers discussed subjectivity in grading so frequently because they were surrounded by it and there was an unspoken implicit acceptance of some degree of non-objective grading. Teachers valued the feedback that grading and assessment have to offer students, but also recognized that grading might not be a reliable, accurate, or “true” measurement of student content knowledge.

Overall, teachers in this study exhibited a lack of content-specific decision making when looking at borderline student work and deciding what grade to assign students. A variety of factors influenced their decision-making, like consequences for the student and student effort.
However, none of the practices hinged on the fact that they taught a particular science course. In contrast, some of their decision-making about grading did seem to be influenced by the fact that they do have a science background; particularly with respect to the scientific principles and understanding of inherent measurement error. The fact that this study only looked at physical science teachers is a limitation, especially with regard to the current claim. This particular group of teachers, who had science backgrounds, used scientific reasoning in their decision-making. A suggested area of future research could look at how teachers of other academic backgrounds make decisions about grading and explore if these teachers utilize scientific or other discipline-specific reasoning in their practices.

The Lack of Discipline-Specific Decision-Making when Deciding to Bump Up

Throughout the course of both sets of interviews, it became evident that the subject matter (physics, chemistry, or earth science) taught by teachers had little to no impact on their decisions about *bumping up* borderline end of term grades. In reviewing the previous quotations in this section, no mention of discipline-specific content knowledge or subject-specific pedagogical knowledge occurred. This is true for the entirety of the interview transcripts. Teachers did not discuss any factors relevant to their decision-making about bumping up grades that were specific to their course content. It would be impossible to distinguish the chemistry teachers from the physics teachers, the physical science teachers, or the earth science teachers, and vice versa. Many of the teachers, as described earlier, often made their final decisions about borderline grades using factors that were *not* related to the specific content they taught; rather they primarily based their decisions on how much effort the student put forth, the student’s circumstances, or the type of grading border.
Making Decisions during the Course. This lack of content-specific decision-making does not mean these teachers did not take content into consideration when making decisions about grading. They often did refer to content related factors, particularly in grading tests. Fern, who mostly taught chemistry classes discussed in detail how certain skills and point allocation could adversely affect a student grade,

I was just talking to a math teacher about this in the hall the other day, that sometimes kids can make many little niggly type tiny, little mistakes, and they can add up to a lower score. But really, they might have used strategies that show that they actually understand the concepts and losing some points for things like significant digits and units, and maybe miscalculating. It's fine to lose some points for those things, because in chemistry details do matter. But to miss a proficiency level, when there's a lot of really great understanding that's being demonstrated…So for that I would make an adjustment. (Fern)

Although this teacher mentioned proficiency and proficiency levels, she does not teach at a proficiency-based school. All of the teachers at public schools in this study were familiar with proficiency-based grading because of L.D.1422. Many schools still assess students for proficiency on standards while still reporting grades using a numerical (0-100) scale. Before entering into this dialog, this participant stated, “Really just the kind of taking a step back and looking at the whole product.” She recognized that consistently repeated small mistakes were not a good indication of the conceptual understanding of a student. This teacher utilized test and assessment retakes, like many of the teachers in this study. Although she did not change individual test grades, she utilized her understanding of chemistry to determine a more holistic and overall picture of what a student was capable of and their proficiency level.
Melissa, physics teacher in this study, recognized a similar situation. She was grading tests when she realized a typically well-performing student ended up getting a lower grade than normal:

I remember scoring a student's test. And I was actually surprised because the student did not do as well as they had in the past... And I just remember being like, what is going on here? I ended up going back because... I make them show all their work. And I could see that he consistently did the same type of problem incorrectly. Like he had forgotten... must have been one of the kinematics equations. He kept forgetting to square something. And he did it consistently. And with my, you know, what do you do in that situation? ... I did still take some points away, but I didn't take all of them every single time.

Both of these teachers made adjustments based on their understanding of the content they were teaching. In chemistry, the teacher believes that details, like units and significant figures matter for her subject area. However, the teachers felt that the number of points taken off for the lack of adherence to these details should not warrant a grade that may indicate a lack of conceptual understanding. The overall conceptual understanding of the student was more important to this teacher than the finer details.

Similarly, Melissa recognized the overall conceptual understanding of the student and realized that the loss of points on a test was coming from one consistent type of mistake. The student continually forgot to square a term in a kinematics equation. Like the chemistry teacher, this teacher chose to modify her point allocation so that the student grade better indicated their conceptual understanding.

These content-related considerations occurred during decision-making about grading before the end of term grades were finalized. As stated in the previous sections, the borderline
grades teachers considered raising were only the quarter, semester, and final course grades –
grades they viewed as more permanent and consequential for students. Despite making decisions
related to the specific content they teach, the specific content taught (earth science, chemistry, or
physics) did not influence if teachers bumped up these borderline grades.

**Making Decisions for End of Term Grades.** When teachers decided to bump students up for
end of term grades, they focused on non-achievement factors that were not related to student
content knowledge. Like many other participants in this study, Fiona focused on non-
achievement factors, especially with students near the pass/fail border. Student effort was
conceptualized and perceived in multiple ways by teacher participants: attendance, participation,
asking questions, wanting to learn, and timeliness. Fiona spoke clearly about raising a student
grade for non-achievement factors “…sometimes if there's like a kid who is at a 68 or 69, and
I've made some kind of deal with them, okay, if you don't have any late assignments, I will make
sure you pass for quarter three.” Again, as previously discussed, all of the teachers raised student
scores that were within a couple points of passing, frequently citing effort.

Student effort was frequently cited as the primary reason for raising student grades that
were within one or two points of a prominent border (pass/fail or A/B). Another factor teachers
referred to was “extenuating circumstances.” This term was used verbatim by multiple teachers.
They often used it to justify bumping up students on the pass/fail border when they were within
five points of passing. When asked about when he would round, Martin stated succinctly: “68,
69. Unless there's some really strange, extenuating circumstances.”

Teachers cared about their students and use their understanding of and perception of
students’ situations to make decisions when grades were within five points of passing. When
asked what factors affect his decision-making Fred spoke more about student situations: “I mean,
certainly a bigger consideration for me is what their situations are. Their situations, not the situation.” Student situations as opposed to the grading situation (deciding on the end of term pass/fail grade) were important to teachers when deciding grades.

Like many teachers, Fern was concerned about consequences for students if they failed. Fern considered multiple aspects of a student’s life rather than only their content knowledge:

“There's like unique circumstances. You know, children that had medical conditions that happened during the year or tragedies, whatever. You don't want to layer tragedies on tragedies.” In Fern’s view, the experience of failing a course is not good for a student and she referred to it as a “tragedy.” Extenuating circumstances can be tragedies and Fern avoids adding more negativity to a student’s life.

At Seacoast High School, the only proficiency-based school in this study, Paul and Piper often teach students in their fourth year of high school. Many of these students have demonstrated proficiency in all of the NGSS Science and Engineering Practices and have passed three science classes and therefore have completed their graduation requirements. Having taught at the school for over two decades, Paul clearly articulated the situation:

I usually don't get kids in physics that are trying for…They have to take three [science classes] and they have to get a [proficiency level of] 3 in the practices. And I rarely have a kid that is doing that. But if the kid came in, and they were attending class and trying and something like that, and then had a 2 on some practice, I would probably give them a 3 at the end. Like, I don't know, if it serves them, you know, if it's helpful at all to retake the course, especially because the system is - there's so many flaws…I would call it compassionate, but somebody might say, lenient or something like that.
Paul cared about his students, like many teachers. Paul engaged in grading behaviors that were representative of a larger theme: teachers will often act with compassion towards their students, especially in instances of potential failure. The participants in this study felt it was uncompassionate and not in the best interest of the student to not bump up grades (for scores that were within five points of passing) for students with extenuating circumstances.

This pattern of behavior speaks to a larger issue in education. These teachers know about their students’ circumstances and situations. Knowledge about students could be gained through various avenues: conversations with students or parents, coaching, supervising extracurricular programs, guidance counselors, or social workers. These teachers knew about their students’ lives outside the classroom. Their knowledge of their students’ situations influenced their grading decisions and behavior. A larger issue may be the lack of similar grade adjustments when students and families that are not able to communicate the existence of extenuating circumstances that impact student performance. Teachers relied on a method of decision-making for the pass/fail border that may not be applied equitably and equally across all students based on a variety of factors.

**Discipline-Specific Decision-Making When Refusing to Bump Up**

Several teachers did not raise grades on the A/B border. Teachers on the A/B border, regardless of school type, often referenced the ability of students to show continued understanding through either school or classroom retake policies. Because students had the ability to improve their understanding and final grade, the teachers who did not raise grades felt comfortable letting the student’s earned grade stand.

Martin felt students had “unbelievable opportunities” at Woodland High School given the school’s policy to allow for retakes for summative assessments. At Marsh Valley, Flynn allowed
quiz retakes as a classroom policy. When discussing whether he rounded up on the A/B border, he stated clearly, “I kind of feel like, you know, there's things they could have done before it got to that point.” On the surface, the statements from Martin and Flynn appear similar to that of Fiona (on page 82) in the previous section. However, rather than solely focusing on “turning things in” the context in which their comments were made focused on students’ academic achievement. They both referenced students demonstrating improved and additional content knowledge.

Fred spoke the most clearly about holding grades right below a mark based on students not achieving the necessary level of understanding. Fred often referenced his commitment to grade students as objectively as possible, based on student content knowledge. With respect to the A/B border, Fred stated,

And so, what I try to do is be really flexible during the quarter, like during the grading period to give them as much help as they need. Give them opportunities as they need to. But then, at the end of the quarter…I'm not willing to like, change the number.

Like many teachers, Fred worked with students throughout the course to help them learn the material. Fred held his stance on not bumping up student grades, despite sometimes feeling conflicted about it, “And remember, just because you have B+ that does not make you a B+ person, right? So, but I still I always question that. I, almost always, every quarter, there's a feeling. That's a bummer, you know?”

However, in conjunction with this comment and previous ones, Fred recognized and acknowledged that a student’s achievement may not meet an “A” level, but he also seemed to place value on some non-achievement factors in his decision-making process. Fred felt he provided students a service or life lesson by not raising their grade based on effort and hoped to
prepare them for future educational experiences. He wanted his students to learn the content and
skills of the course and keep improving their understanding. Fred also wanted students to use
grades as feedback for their learning and feel proud of their accomplishments throughout the
course, rather than just satisfied with their grades.

In the instances where teachers did not raise grades, they focused on student content
knowledge or the lack thereof as a reason to not bump up end of term grades. However, student
ability to demonstrate a certain level of content knowledge often incorporates some level of non-
achievement factors, like effort. To demonstrate a certain level of content knowledge (despite not
at an earlier stage), students often must continue learning the material and engage with whatever
retake policy exists.

At Woodland High School, Mary decided not to student pass a student for the course
after he failed the final exam. When speaking about the chemistry student, Mary said,

I had a kid that got a 70 and got a 72. And then on this final exam, he got a 17. That was
not going to be a passing. So what that tells me is that the 70 and the 72 were gifts
already because he only got a 17 on the final.

Mary decided to hold the end of term student grade as is (per the numeric calculation on
Infinite Campus). Mary justified this by referencing the student’s demonstration of content
knowledge. Although the student earned passing grades during each semester or quarter, the final
exam demonstrated a profound lack of understanding. Mary valued the school policy of
emphasizing summative assessments as a measure of student understanding. The low score on
the final demonstrated a low understanding of content; therefore, she felt justified keeping the
failing course grade.
Teachers also talked about instances in which they felt pressured to change an end of term grade but ultimately did not; regardless of what border they fell on. The specific pressure teachers felt was unclear and not a focus of this research. However, during the interviews, teachers did express feeling tension or external pressure to have a student grade be at a certain level.

Fern explained and articulated her dilemmas in negotiating grading decisions. During the last part of the school year, Fern had a student who needed to take a test (due to absence) and had several late assignments. The grading of assignments was time-sensitive due to the terms of a scholarship the student was applying for. Fern was given little notice. Fern struggled with whether to grade the missing assignments and consulted with her colleagues. Ultimately, she decided to grade those missing assignments. She recounted her experience while grading,

In the meantime, I scored the exam that the child took, and it was very bad, very, very bad. And so when I put the exam in, and whatever they got on these two little practice things, it was unlikely that their new score was going to be substantially better. So we decided that the best thing to do was just score the two things, even though it like goes against the policy of what I do, and whatever. And I didn't do it for other kids. Score the two things. That way it's not me holding out on the child being able to have the scholarship. It's just based on the kid’s performance. So I scored the two things that were not very good. This test was terrible. I put in all the grades, and I replied to them, all the grades are done. I scored everything. And then the kid called me back like 30 minutes later: “Is there anything else I can do? I really need my score to be higher…” I replied, with the mom cced [that] I'd already made an exception scoring the things that were late for March. And I didn't do that for anybody else. And that alone makes me
uncomfortable. And I cannot give you another assignment to do that I'm not going to offer to everybody else. And there isn't even time for you to do an assignment that's meaningful and score it. So no, I'm sorry. This is what it is. And she said: “Okay, thank you.” And then I didn't hear anything else. So sometimes I can push back.

Teachers will hold grades, and not raise them, based on their assessment of a student’s effort and/or content knowledge. The statements presented do not reference specific content knowledge but refer generally to student understanding. Despite having particular content goals throughout the course, end of term grades are comprised of many individual content goals tied together with overarching themes. Therefore, the lack of discussion of specific content pieces is not surprising. When deciding to hold a student’s earned grade, teachers spoke vaguely about their students’ content knowledge when deliberating about end of term borderline grades. Teachers considered their holistic understanding of the student and then made a decision.

Patterns of Behavior

In the previous sections, teachers’ use or non-use of subject specific content knowledge was examined with respect to their decision-making processes in grading. In deciding to hold a borderline student grade below a mark, teachers referenced a perceived lack of demonstration or possession of content knowledge. The lack of academic achievement on the student’s part occurred throughout the course, rather than on individual assignments. The students who did not have their borderline end of term grades raised were those who demonstrated a consistent lack of content knowledge to achieve certain marks.

However, when making decisions about whether to raise end of term grades that were just below (within one or two points) of the pass/fail or A/B border, teacher perception of student effort was the primary consideration. When course grades dropped below five points of passing,
teachers often used their understanding (and empathy) of individual student circumstances and situations. This type of decision making was only with respect to the pass/fail border and not any other grading border.

When deciding borderline grades, teachers considered multiple factors, regardless of their school grading structure. Despite utilizing their understanding of discipline-specific content knowledge throughout classroom assessment and grading, teachers only partially rely on this content knowledge when deciding end of term grades. Teachers felt some pressure, either external or internal, when making decisions about grades. Content knowledge was used as a justification to hold student grades as is rather than bumping up to the next mark.

In making the decision to bump student grades up, teachers often justified this through non-academic student attributes: effort, which students have some control over or “extenuating circumstances”, over which students have no control. When student grades fell within a close range, one to two points below a border, teachers often felt like students were working as hard as they could or they were not. That perception determined whether to raise student grades on the pass/fail and A/B border.

When student grades are at least five points or more below the border for passing, teachers often considered the student’s circumstances. These circumstances were described broadly in the data: student home life, personal or family medical conditions or illness or loss of loved ones.

**Schema Switching.** Teachers treated their decision making about a grade differently based on different situations or factors. Said another way, teachers frame certain grading activities differently than others. Throughout a course, teachers did not worry about borderline grades. As stated before, many teachers employed retake policies in their classrooms to encourage students
to keep learning and ultimately improve their end of term grades. However, when deciding end of term grades, teachers treated this process differently than classroom grading. Different borders mattered more than others and teacher decision making often utilizes non-achievement factors. Additionally, teachers treated different borders, different courses, and sometimes different students differently. This behavior speaks to a variation in underlying epistemic framing of grading, which will be discussed in the next chapter.

**Summary**

In summary, *end of term* borderline grades were the grades teachers focused on, deliberated about, and ultimately made difficult decisions about. Consistent with previous research, teachers’ concerns about the negative and positive consequences of numerical grades factored into and motivated their decisions about grading. Only certain borders for grades warranted a level of decision-making or adjustment that teachers were able to and wanted to describe. The pass/fail border was universally discussed when describing experiences surrounding borderline grade. The A/B border provided challenges or decisions to be made for some, but not all, teachers.

The themes of inaccuracy, bias, and subjectivity were often areas of concern for teachers. Additionally, teachers recognized the multiple areas where inaccuracy and bias could appear in their practices. Most teachers generally accepted that different aspects of grading are inherently subjective and have some degree of error.

Teacher content knowledge also affected their decision making differently for end of term borderline grades. Teachers often bumped students up in grades based their perceptions of student effort or personal circumstances. Students’ content knowledge did not significantly influence teachers’ decisions about whether to bump up. However, content knowledge did factor
into decisions not to raise grades. In justifying their decisions students’ earned end of term grades stand, teachers often cited students’ lack of content knowledge.

Lastly, this pattern of decision making could be indicative of larger behaviors in how teachers frame grading in different situations. Classroom assignment grading and end of term grading are not the same process and have different perceived stakes or consequences for students. In addition, not all end of term grading decisions were perceived similarly. Different factors seem to have varying levels of influence over teachers’ grading decisions in different situations.

In the next chapter, the larger patterns of teachers’ decision making will be explored further with respect to prior research. Additionally, these patterns will be explored using a lens of epistemic framing and schema switching.
CHAPTER 6

DISCUSSION OF FINDINGS AND CONCLUSION

This chapter will address this study’s findings in relation to my research questions and prior research on grading. After revisiting the research questions, the key findings of Chapter 4 and 5 will be briefly summarized. These key findings will then be explored with a focus on examining areas of consistency and inconsistency with research on teachers’ grading practices. Patterns that emerged in teacher practice will then be broadly explored using the theoretical concept of framing. Finally, suggestions for future research and implications for policy and practice will be discussed.

The purpose of this phenomenological study was to describe what factors influence how high school physical science teachers determine borderline grades and how these decisions are made. This research was guided by one broad research question with two sub-questions, which are restated here:

How do teachers experience grading in borderline situations and make decisions about borderline grades?

a. In what grading context and school setting do teachers make these decisions?

b. What factors influence their grading decisions and how?

In order to answer these questions, twelve high school physical science teachers in Maine were interviewed during the 2020-2021 school year. Two semi-structured interviews were conducted with all twelve participants and transcript data were analyzed using an interpretivist approach to explore how they experienced the task of determining borderline grades.
Review of Broad Findings

In chapter four, after describing the state context, teachers’ perceptions of their school or district grading polices and expectations were explored as well as teachers’ self-reported experiences while grading. Teachers’ experiences were then used to characterize three different types of high school grading systems: proficiency-based, systems with many grading guidelines and systems with fewer grading guidelines. All of the schools except one used the standard 0-100 numerical scale for reporting classroom and end-of-term grades. The majority of the schools in this study were identified as having fewer grading guidelines with teachers reporting feelings of autonomy and freedom in their decision making about grading practices. Despite differences in schools’ approaches to grading and grading policies, teachers were often able to exercise individual choice, especially when making decisions about borderline grades.

In chapter 5, factors that influenced teacher decision making about borderline grades were discussed. Regardless of what types of expectations or policies existed at their different schools, the teachers in this study engaged in a similar process of decision making and grading practices.

Teachers only considered end-of-term grades when making decisions about borderline grades. They did not deliberate about grades on classroom assignments like tests, quizzes, or lab reports regardless of how close to the border those grades fell. Retake policies, implemented by the individual teachers or by the school, seemed to alleviate teacher concerns in deciding to leave classroom grades as is.

For end-of-term borderline grades, teachers only engaged in the process of increasing scores slightly below the border – typically within 1-2 points, but sometimes within 5 points of passing. Teachers in this study were adamant about not lowering the scores of borderline grades.
that were slightly above a border. This practice implied that there may be cultural norms surrounding end-of-term grades, specifically with respect to the access parents and students had to grades through the online grading platforms. It is possible that teachers and other stakeholders accept the end-of-term grade calculated by the grading software as a minimum earned score with opportunities for it to increase.

The decision to increase a student’s end-of-term grade was most heavily influenced by the teacher’s perceptions of the student’s effort. However, some teachers consistently raised student grades that were 1-2 points below a border regardless of student effort. For grades that were further below a border (more than 5 points), teachers often would not increase or modify the grade. However, there were cases described by the teachers as students with “extenuating circumstances” where teachers would raise the grade.

Different borders were more or less challenging to assign grades to. All of the teachers in this study found it very challenging to assign failing grades to students that were within 1-2 points of the pass/fail border. Teachers often raised these grades hoping to minimize the negative consequences they perceived for students. Along the pass/fail border, teachers were concerned about the inability of students to receive course credit and/or graduate or being required to take summer school. The other prominent border was the A/B border. Approximately half of the teachers increased grades below the A border based on their perceptions of students’ effort. This choice acted as a reward mechanism, but for some teachers it helped avoid perceived potential negative consequences for students. Along the A/B border, negative consequences included not making “High Honors” and potentially damaging college acceptances. No other grade or proficiency borders presented significant challenges to teachers.
In deciding to increase borderline end-of-term grades, teachers often referenced non-achievement factors in their decision making, rather than discipline-specific content knowledge. In deciding to hold grades as is, however, they would often justify this choice through the student’s inability to demonstrate a necessary level of content understanding.

The secondary science teachers in this study often referenced the subjectivity, inaccuracy, and bias in grading. They viewed their grades as having a margin of error. This understanding of inherent error was often used in their reasoning, sometimes in conjunction with student effort, to justify increasing end-of-term grades. It is possible that because these teachers have backgrounds in science and scientific measurement, the notion of measurement error in grading influences their grading decisions. This is an area of limitation for this study, but also a potential area of future research.

Many of these findings from this study are consistent with previous research on grading and determining borderline grades, which will be discussed in this chapter. However, this study also indicates some potential gaps in our understanding of teacher practices surrounding borderline grades. Inconsistencies with prior literature will also be discussed in this chapter.

In this chapter, teachers’ borderline grading practices are explored using the psychological concepts of frames, framing, and schemas (Hammer et al. 2005; Hammer et al, 2006; Hammer & Elby, 2010; Piaget, 1971; Tannen, 1978, 1979, 1993). These grading practices are connected to prior literature on assessment and grading, specifically focusing on the push from educational and grading researchers to grade for students’ summative academic achievement rather than other factors (Brookhart, 2003, 2004; Cizek, 1996; Feldman, 2018; Guskey, 2011; Marzano, 2000; Snowman & Biehler, 2003). The use of an interpretive lens in
this study provides deeper insight into the question of why, despite professional development, teacher practices are slow to change.

**Connections with Prior Research**

As stated earlier, the findings from this research present both similarities and differences with prior literature. The vast majority of research on grading has been performed quantitatively through large surveys of teachers of all disciplines across varying grade levels (Cross & Frary, 1999; Guskey & Link, 2019; McMillan, 2001; Randall & Engelhard, 2010). Given this qualitative study’s focus on high school physical science teachers, the findings confirm broader trends found elsewhere in the literature, but also highlight potential areas of practices specific to this population subset. Several specific areas of findings will be discussed in relation to previous work on grading and borderline grades: the borders that teachers pay attention to, teachers’ concerns with error and inaccuracy in grading, and the practice of increasing borderline grades and the non-achievement factors that influence teachers’ decision-making.

**Important Borders in Borderline Grading**

The teacher participants in this study universally reported bumping up students’ end-of-term grades along the pass/fail border and sometimes engaged in raising grades along the A/B border as well. The teachers in this study did not give equal attention to the D/C or C/B borders in their decision making and often found these borderline grades easy to determine.

Although the treatment of pass/fail and A/B grades appears to be consistent with previous research, the lack of attention to the B/C and C/D border is not. These “middle of the range” borders did not warrant the same level of consideration as borders on either end of the grading spectrum. This practice contradicts some prior research.
Randall and Engelhard (2010) found that K-12 teachers of various subjects would often raise hypothetical students’ borderline grades for all grading borders if they were right below the border for a mark. This was true of each grading border: pass/fail, C/D, B/C, and A/B with the largest effect being on the pass/fail border. Other research has found that K-12 teachers often raised the grades of failing students who exhibited favorable non-achievement characteristics (Brookhart, 2003). Brookhart’s results agree with teacher decision-making on the pass/fail border for this current study.

The two previously mentioned studies focused on a mix of K-12 teacher practices rather than high school physical science teachers specifically. It is possible that secondary teachers make different grading decisions than elementary grade teachers.

Another possible reason for this variation in practice might be because of changes in the grading system many of these teachers used. Webb et al. (2020) found that the type of grading scales used in introductory college physics courses strongly affected the proportion of these middle of the range grades. Over the course of a ten-year study, physics instructors used both numerical 0-100 percentage scales and 0-4 scale (where each number corresponded to a letter grade. When instructors used the numerical percentage scale, they were far more likely to assign grades less than a C-. This suggests that differences in student grades, specifically in the middle to lower range, might be a result of the type of grading system used rather than teacher variation.

The lack of attention to the C/D and C/B border by the teachers in this study might support the findings of Webb et al. (2020). The numerical 0-100 percentage scale might increase the possibility that these middle to low scoring students are being overlooked. Paul and Webb (2022) also found that switching from a percentage-based scale to a 0-4 scale could decrease grading inequities by up to 25%.
In the next sections additional teacher concerns, experiences and practices will be discussed. The teachers in this study expressed concerns about error in their grading which manifested differently at different borders. Additionally, non-achievement factors had varied influence on the different grading borders.

**Concerns with Error in Grading**

An emergent theme from this study was teachers’ shared concerns with error and inaccuracy inherent within their classroom grading. The teachers in this study often referenced the idea of unreliability in their end-of-term grading, citing the accumulation of inaccuracy over the term and/or bias or subjectivity in *what* they decided to grade and *how* they decided to grade it. For example, teachers in this study recognized that the points allocation they used for individual test questions, tests, and lab reports possessed a certain degree of arbitrariness. They also recognized they chose what student work to grade and how to grade it which also contributed to subjectivity and bias; they graded what they felt was important. Teachers felt that projects, journals, and assignments graded with new or developing rubrics were subjective. Furthermore, when grading assignments, teachers also recognized they might misallocate points or incorrectly add up points which would lead to inaccuracy in the end-of-term grade. Overall, this aspect of grading was viewed negatively by the teachers in this study and often provided them with motivation to adjust end-of-term grades.

This particular aspect of this study presents an interesting contrast with prior research. Prior research has focused on the inaccuracy, bias, and subjectivity *researchers* perceive in teacher practices (Brookhart, 1993; Kunnath, 2007), rather than how *teachers* perceive this aspect of grading. In the literature, researchers overwhelming view grades as a measurement of students’ academic achievement and teachers’ grading practices have been criticized as being
subjective and inaccurate. Very few studies have looked at teachers’ perceptions of the inaccuracy, bias, or subjectivity present in their grading practices, why they might engage in these processes, or how they view it.

Inaccuracy and subjectivity in K-12 teachers’ grading practices were thought to be due to a lack of training in educational measurement (McMillan, 2003; Stiggins, 1999). However, Brookhart (1993) found that training in educational measurement had little effect on K-12 teachers’ grading practices, specifically with respect to increasing a numerical grade’s accuracy or validity in measuring academic achievement. Teachers’ lack of knowledge about inaccuracy, bias, or subjectivity in grading may not be the primary factor that determines how and if teachers do engage in questionable grading practices.

Some prior studies pointed out that teachers possess an awareness of their ability to engage in inaccurate grading practices. Zoeckler (2007) found that secondary English teachers often used the weighting of grades to emphasize important aspects of the course or teachers’ own personal philosophies. During interviews, teachers in this study expressed awareness of their ability to modify assignment weights to emphasize self-selected areas of importance or to reduce potential student failures by increasing the point allocation for assignments that students performed “well” on and, vice versa, by decreasing point allocations for assignments students did not perform well on.

Zoeckler presents similar findings to the current study: teachers recognized they could adjust grades to place value on certain aspects of the course. However, Zoeckler did not examine how teachers perceived this ability and practice. If the majority of researchers agree that inaccuracy, bias, and subjectivity in grading should be minimized, it would be worthwhile and informative to examine how teachers perceive these practices through further research.
Error Concerns Manifest Differently at Different Borders. An interesting feature of teachers’ considerations about inaccuracy and error in end-of-term grades in the current study was that this concern was relevant for different borders preferentially. Teachers often leaned heavily on this reasoning (that the computer calculated end-of-term score was a range rather than a point) in making decisions about the pass/fail border. At this border, teachers often gave students the final end-of-term grade that was at the high end of the score range they felt existed within their grading.

However, on other borders, this facet of reasoning was less utilized or not used at all. The C/D and B/C borders did not prompt any error-based reasoning. The A/B border did, however, cause some teachers to consider error in their grading. Again, only half of the teachers rounded up on the A/B border, with the other half holding the score as is. For the teachers that did round, many referenced concerns with error or inaccuracy in their end-of-term grades, in addition to other types of reasoning.

All the teachers in this study shared their reasoning about inaccuracy and error in grading, but they did not apply this reasoning universally across all borders. Different borders were treated differently, which will be discussed in more detail in the next section.

Increasing Borderline Grades for Non-Achievement Factors

Teachers only engaged in the process of bumping up end-of-term grades (rather than bumping them down) or decided not to raise the term score. This decision-making process was explored in the previous chapter. The practice of raising end-of-term grades (rather than lowering them) is consistent with previous research. Brookhart (1993) found that K-12 teachers would engage in numerically raising end-of-term grades for non-achievement factors even when they had formal instruction in educational measurement. Bumping up grades was not a universal
practice in that quantitative study however. Much like the teachers in this study, Brookhart found through surveys that teachers would keep students’ end-of-term grades as is if they decided not to raise them, although there were only a few scenarios in which teachers could opt to lower student grades (which a small proportion did).

Randall and Engelhard (2010) found a similar pattern of K-12 teachers raising students’ report card grades for non-achievement factors. This was especially prominent for “low ability” students. In a quantitative survey, hypothetical, low-achieving students, particularly those along the pass/fail border, who demonstrated high effort and excellent behavior, often received a significant bump in their final report card grades. Brookhart (2013) similarly found that non-achievement factors (such as behavior, academic effort, and ability) have a larger positive impact on the grades of lower performing students.

Teachers’ reluctance to assign a failing grade earned by a student at the end of a course might have larger implications for the American educational system. For less explored reasons, teachers more readily pass failing students and are less likely to increase the grades of already passing students. Students who demonstrate certain non-achievement characteristics (i.e., behave in a way deemed “appropriate” by teachers in school and/or show effort to learn the content), but demonstrate a lack a sufficient level of content-knowledge, can often still earn passing course grades.

**The Mixed Influence of Non-Achievement Factors**

The teachers in this study considered non-achievement factors to varying degrees in their decision making. The utilization of non-achievement factors by teachers in this study to determine end-of-term borderline grades is consistent with prior literature. In the next few sections, commonly cited non-achievement factors will be discussed with respect to the current
study. Teachers’ self-described decision-making will be examined using three well-documented non-achievement factors: effort, behavior, and ability cited in several prior studies (Brookhart, 1993; Cross & Frary, 1999; Kunnath, 2017; Randall & Engelhard, 2010). Additionally, other non-achievement factors teachers identified will also be compared to prior research on grading.

**Effort, Behavior, and Ability.** A student’s academic effort in a class was conceptualized broadly by the teachers in the current study. Teachers in this study considered the following student characteristics to be indicative of their academic effort: timeliness of completion of work, being focused on learning during class and taking advantage of classroom opportunities to earn a good grade (e.g., test retakes or participation credit). Additionally, teachers sometimes considered how willing a student was to ask for help in class or come in for extra help. Teachers were very specific about the academic behaviors they were looking for in students given the structure of their classroom and their classroom grading.

Although teachers in the current study often referenced certain kinds of student behavior they considered in grading decisions, they only described behavior that related to academic effort. They did not talk about other kinds of non-academic or normative student behavior in the classroom. None of the teachers in this study referenced specific behaviors about students they found disruptive, nor did they consider these behaviors in the determination of the final grade. Although they sometimes referenced whether they “liked a kid or not”, they strongly professed to not using this attribute to determine the end-of-term borderline grade.

Student ability was not referenced by the teachers in this study when they described their grading practices and borderline decision-making. Although they described accommodating differences in student readiness or learning style during instruction, the teachers did not describe ability-related factors appeared in their grading rationale.
Randall and Engelhard (2010) described these three common non-achievement factors that influenced grading decisions generally in their survey of K-12 teachers. In their study, the concept of effort was a broadly described term, i.e., a student is working hard on their academics. Behavior was not clearly defined by the authors who presented this concept as a dichotomy: either excellent or poor classroom behavior. Excellent behavior was described as students exhibiting good manners and not talking out of turn, while poor classroom behavior was described as the opposite of those behaviors and also being disobedient in class.

By contrast, Randall and Engelhard (2010) described how teachers in their study considered student behaviors that would be disruptive to a classroom learning environment in their grading decisions. In the 2010 study, behavior had the largest effect on teachers’ grading decisions, outweighing effort. Behaviors like talking out of turn and being disobedient often resulted in teachers lowering students’ borderline report card grades or holding the grade as is (rather than increasing it to the next level). Randall and Engelhard (2010) present what might be an antiquated vision of student behavior in the classroom. Teachers in this study never mentioned students’ classroom manners or any expectations about when students were to talk in class. Students that exhibit poor behavior might be more challenging for teachers to “like.” It might be possible that teachers used “liking a student” to refer to student behavior rather than citing instances of behavior directly. The participants in the current study adamantly and continually stated they focused on not factoring whether they liked a student.

The one behavior that teachers did refer to specifically and negatively in the current study was cheating. There were several instances where teachers would lower scores (or give zeros) in classroom grading. Additionally, they also relied heavily on the summative assessments (quizzes, tests, and exams) to gauge the achievement of a student suspected of cheating. However, if a
student was suspected of cheating, they would sometimes modify classroom grades to reflect the behavior, often giving zeros, but left the end-of-term grade as is.

It is possible that student behavior has a lower impact on grading decisions for high school teachers than for teachers of lower grades. The teachers in this study worked with students in grades 9-12, with a slight skew towards the upper end of the range (based on the science subjects they taught). Although poor behavior does occur in high school classrooms, it might be less prominent at these upper grades and in more academically advanced courses or tracks. Behavior might be a more prominent factor for elementary or middle school teachers making decisions about borderline grades and therefore, “poor” behavior might have more implications for learning and assessment. When students reach high school, they typically have better self-regulation skills.

Additionally, middle and high school are typically when students have more independence in deciding whether to attend school or not, and truancy becomes an issue (Alexander et al., 1997). It could be possible that student absenteeism leads to a drop in instances of unproductive student behavior in the classroom. Teachers in this study often talked about the importance of having students attend school and often were frustrated when students were failing due to absence. They also referenced absenteeism as a concern in determining borderline grades. They found it difficult to pass students that did not come to school and therefore did not show the requisite academic effort to achieve the desired level of content understanding. However, they also found it acceptable to pass students that did come to school and did not achieve the desired level of content understanding, despite their grades being close to the numerical or proficiency cut-offs.
The focus of this study was on borderline grades, which emerged as only having significant relevance for end-of-term grades rather than classroom assignment grades. Teachers in this study purposefully ignored certain kinds of student behavior in their final grade determination, focusing more on academic achievement, effort and some additional factors that will be discussed later.

**The Non-Uniform Influence of Effort on Decision-Making at Different Borders.** Like teachers’ concerns about the inaccuracy and error in grading, teachers’ perceptions of students’ academic effort affected different borders differently in the current study. All of the teachers in this study were willing to bump students up on the pass/fail for effort if student grades fell within a few points of passing. However, along the A/B border, some teachers were unwilling to factor effort into their decision-making about borderline grades and held students’ end-of-term grades as is.

Student academic effort was a facet of teachers’ reasoning in making decisions about end-of-term borderline grades. However, it was not applied consistently across the two important grade borders and was not applied to the “middle” borders, C/D and C/B, of borderline grades. Previous research supports the finding that student effort has a large influence on teachers’ final grades. Further, prior research shows that lower achieving students often received larger grade increases compared to higher achieving students (Brookhart, 1993; Randall & Engelhard, 2010). Additionally, in prior studies, students’ academic effort affected all grading borders to a varying degree, which contradicts with the current study’s findings.

To summarize, consistent with previous research, students’ academic effort was a strong non-achievement factor in teacher decision-making for borderline grades. However, effort did not have a consistent degree of influence for all grading borders. Additionally, student behavior
appeared to have no effect on teachers’ decision making in the current study, despite being documented as having significant influence in prior research (Randall & Engelhard, 2010).

In the next few sections, a third well-documented non-achievement factor, ability, will be discussed with respect to the current study’s findings. Additionally, two potentially unexplored non-achievement factors will be discussed in subsequent sections: student’s personal extenuating circumstances and teachers’ desire to avoid negative consequences for students.

**Ability not cited as a factor.** Teachers in this study did not articulate any reasoning that factored students’ academic ability to determine end-of-term borderline grades. Teachers often referenced students’ effort and willingness to learn the material despite obstacles, but never used their perceptions of students’ ability to make end-of-term grade determinations.

In prior literature, ability has been defined as teachers’ perceptions about what students could achieve academically, determined through classroom observations or standardized test scores (Cross & Frary, 1999; Randall & Engelhard, 2010).

A study of 9-12 teachers of all subjects found that ability was *not* a factor in high school teachers’ grade decision-making (McLean, 2018). This supports the current study’s findings but contradicts a significant body of prior research.

Previous literature has cited K-12 teachers’ perceptions of student ability as a significant factor that influences grading decisions (Cross & Frary, 1999; Randall & Engelhard, 2010). In a study of middle and high school teachers and students, Cross and Frary (1999) found that approximately 50% of the teachers in their study considered ability when determining a student’s grade. When Randall and Engelhard (2010) examined K-12 teachers’ report card grading practices, teachers did consider a student’s ability level. In that study, teachers’ consideration of student ability had a larger impact when considered with both achievement and non-achievement...
factors (e.g., behavior and effort). Additionally, lower performing students (of low ability) tended to get larger increases in grades than did higher ability students.

There may be several possible reasons for contradictions in how teachers weigh and use student ability in determining grades. Prior studies that suggested ability was a substantial grade-determining factor were conducted one to two decades ago and it is possible changes in teacher practice have occurred. Additionally, high school teachers might approach grading differently than those of younger grades. Many of the studies referenced above survey K-12 teachers rather than 9-12, specifically. It is also possible that a shift in grading practices has occurred. Teachers’ approach to grading might be more on student learning rather than their perception of what a student is capable of learning. Additionally, in the high school context, student ability might be closely tied to students’ academic achievement. Grades, if viewed objectively, reflect students’ academic ability and achievement. Therefore, ability may not be viewed as a factor on its own and grading adjustments are made based on other, more distinct factors.

**Less Studied Non-Achievement Factors.** In the current study, teachers used additional reasoning to make decisions about end-of-term borderline grades that appear to be less well described in existing literature. In addition to effort, they often relied on their understanding of students’ personal circumstances and their perception of negative consequences that could affect a student as a result of the end-of-term grades. These extenuating circumstances are described in the following section.

*Extenuating circumstances as a non-achievement factor.* Many teachers in this study cited a student’s personal, extenuating circumstances as a factor that influenced their determination (and bumping up) of borderline grades. This was particularly prominent at the pass/fail border. Teachers often bumped up failing student grades in an act of compassion and concern with
students’ less than ideal circumstances. Commonly mentioned examples of extenuating circumstances were student illness or the death of a loved one or family member. Teachers recognized that students may have unequal opportunities to learn content and/or demonstrate their content understanding.

In this study, the influence of extenuating circumstances most often affected failing students. This has several important implications. Teachers may or may not actually know the personal background and circumstances of all their students to the same level of depth. Based on several factors, some of which could be cultural, teachers might have more or less insight into the lives of different students. When making decisions based on perceptions of extenuating factors, this non-achievement factor may be applied preferentially to students or families that communicate this. If a student, parent, or family lacks this communication, a hypothetical failing student might not get a grade raised.

Teachers’ perceptions of students’ personal circumstances have not been explored thoroughly in the literature. Prior research has shown that high school teachers often want to facilitate student success and in doing so will sometimes consider “exceptional circumstances” in making general grading decisions (Kunnath, 2017). However, teacher grading practices and decision making that considers “extenuating” or “exceptional” circumstances have been largely understudied.

Avoidance of Perceived Negative Consequences as a Non-achievement Factor. The teachers in this study considered the consequences of failing a course or having to go to summer school as negative consequences. This reasoning appeared both along the pass/fail and A/B border. Being promoted to a subsequent course or grade level without demonstrating proficiency
or graduating without necessary content understanding was not perceived as a negative consequence.

In choosing to avoid more immediate negative consequences for the student, this pattern of grading behavior could be contributing to a lack of readiness or achievement later in high school or beyond. Teachers may be promoting students who have not met certain standards (content of proficiency), thereby weakening the value and accomplishment behind earning a high school diploma.

The avoidance of negative consequences for students has not been studied as a factor itself in previous research. However, teachers have often been found to want students to be successful which can influence their grading decisions (Kunnath, 2017; McMillan & Nash, 2000; Tierney, 2015). Additionally, teachers’ grading rationale was found to be influenced by the future opportunities that might be given or taken away from students in a study of high school English teachers (Tierney, 2015).

In trying to promote student success, teachers might use grading to mitigate consequences that might negatively impact students. Brookhart (1993) and Bishop (1992) found that teachers could not be both advocates and judges for their students and their academic performance. The two roles were incompatible to maintain in classroom grading practice. The results from this study indicate that these conflicting roles still exist in teacher grading practices, specifically with respect to borderline end-of-term grades.

**Understanding Borderline Grading Practices**

The patterns of grading behavior from this study indicate that teachers use a variety of factors, both achievement and non-achievement factors, to determine borderline grades. This is consistent with the previous literature on classroom grading practices. However, some of the
non-achievement factors, primarily effort, were more prominent in this study than other non-achievement factors related to students’ normative behavior or ability found in other studies. Furthermore, high school teachers’ consideration of “extenuating circumstances” and avoidance of negative consequences when making decisions about borderline grades warrants further exploration and research. Figure 6.1 below shows the non-academic factors that teachers in this study cited as influencing their borderline grading decisions.

Figure 6.1 Prominent Non-Achievement Factors in Teachers' Decision Making

<table>
<thead>
<tr>
<th>Factors that Influence Decision-Making in Borderline Grading Situations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time of Year</td>
</tr>
<tr>
<td>Grading Border</td>
</tr>
<tr>
<td>Perceived Student Effort</td>
</tr>
<tr>
<td>Perceptions of Student’s Extenuating Circumstances</td>
</tr>
<tr>
<td>Avoidance of Potential Negative Consequences</td>
</tr>
<tr>
<td>Understanding of Underlying Error and Inaccuracy in Classroom Grading</td>
</tr>
</tbody>
</table>

It is possible that secondary school physical science teachers vary in how they consider non-achievement factors when determining grades. Some research has looked at the grading practices of teachers of different science disciplines (Duncan and Noonan, 2007; Feldman, et al., 1998). In studies of science teachers, it was found that teachers vary in what types of assignments they grade, but remained relatively similar in their end-of-term grading practices.
Different non-achievement factors might play more or less of a role in secondary school grading as compared to earlier grade levels.

Consistent with previous literature, findings from this study suggest that teachers do consider non-achievement factors when determining end-of-term borderline grades. Although this practice might seem “hodgepodge” (Brookhart, 2001; Cizek et al., 1995; Cross & Frary, 1999; McMillan, 2003), I would argue that it is part of a much larger pattern in teachers’ approaches to grading, which I will call “schema switching.” When teachers approach grading, they do not approach it as a single, clearly defined activity; rather “grading” subsumes a variety of different socially and cognitively situated behaviors. In order to introduce this idea, frames and framing will be discussed in the following section before discussing schema switching in more detail as it applies to this study.

**Decisions about Borderline Grades as Schema Switching**

Based on the key findings described from this study, I hypothesize that teachers do not approach grading as a single activity. Rather, grading, and borderline grading, specifically, are experienced or approached through a set of frames or schemas, with different grading decisions essentially being thought of as different contexts and thereby framed and reasoned through differently.

Framing, as discussed in Chapter 2, is a preferential way of noticing features of a situation and then engaging or reacting to the situation based on particular elements of reasoning with different framing resulting in different patterns of behavior.

In my analysis, I draw on Tannen’s work to provide a conceptual underpinning to interpret teacher grading practices in a novel way, allowing a deeper understanding of their
grading experiences, perceptions and decision-making in grading; however, I did not employ the rigorous level of discourse analysis used in Tannen’s work.

Assessment and educational measurement experts argue that grading should serve as an objective measure of student achievement and be gathered through reliable and valid classroom assessments (Allen, 2005; Brookhart, 1993; Guskey, 1996; Snowman & Biehler, 2003; Stiggins, 2001). This view portrays grading as a single frame or schema. In this interpretation, the single, salient factor for grading decisions is the demonstration of student content knowledge or skills. A student’s classroom or end-of-term grade hinges on this single factor: the demonstrated acquisition of specific academic content knowledge. In subsequent sections, I will refer to this as a content-focused schema for grading. By contrast, teachers in this study used this schema to frame certain grading activities, but not others. These differences in practice will be described in the following sections. Figure 6.1 presents my revised conceptual model depicting the factors teachers’ cited in this study as influencing their decisions about students’ borderline grades.
Content-Focused Schemas

Although this study did not focus on classroom grading decisions for specific classroom assignments, teachers often referenced student content knowledge and skills they needed to assess and grade. Their comments also revealed that participation, effort, and completion of work did factor into classroom grades, however; student grades were based primarily on academic achievement, which is consistent with prior research (Guskey & Link, 2019; Randall & Engelhard, 2010).

Teacher practices around borderline grades in the current study were only relevant for end-of-term grades. Teachers did sometimes utilize a content-focused schema for grading throughout the course and in some cases when determining borderline grades. It appears that classroom assignments were easier to make decisions about. Teachers issued classroom grades to students without making adjustments. Adjustments were only made during the grading process.
when teachers tried to grade more accurately for content knowledge and skills on particular assignments. Borderline grades in classroom grading were likely made within a content-focused schema and utilized academic achievement as the driving decision making factor.

The teachers in this study who worked at schools that used numerical grading applied a content-focused schema for borderline grades on the C/D and B/C borders. Additionally, half of the teachers in this study determined end-of-term A/B grades using students’ content knowledge and academic achievement. The numerical course average was determined mostly from students’ content knowledge, the most salient feature of this decision-making process. This approach to grading was not true of all grading borders.

**Alternate Schemas for Grading**

Teachers engaged in different decision-making at different end-of-term grading borders, suggesting that they frame these situations differently than the other grading situations previously described. When deciding about end-of-term grades, the teachers in this study utilized other aspects of their own knowledge or their perception of the student rather than solely relying on students’ content knowledge.

With all end-of-term grades, teachers had the choice to post the grade as is or choose to modify it (bump up). Teachers felt that different situations warranted different considerations. Different features of students’ attributes or personal situations varied in their importance when teachers made their grading decisions. Teachers’ variability in what classroom and student aspects they paid attention to and used to make grading decisions strongly suggests their use of alternate frames or schemas for some borderline grades. Many factors identified in this study had varied significance and impact on teachers’ end-of-term grading decisions, depending on the border.
The time of year was the primary factor teachers considered, hence the focus on teachers’ end-of-term grading practices. The following additional factors were identified as being meaningful to teachers in determining end-of-term borderline grades: the particular border the grade was on, their perception of the effort the student put forth (or the effortful behavior the student exhibited), their perception of the students’ extenuating circumstances, their avoidance of potential negative consequences for the student, and their own understanding and acceptance of the error, inaccuracy, subjectivity, and bias in their grading.

Teachers relied on content-driven schemas for decision-making along the C/D and B/C borders. Approximately half of the teachers used this grading decisions along the A/B border. Three alternate schemas for grading were present along the A/B border and two along the Pass/Fail border. These will be referred to as alternate schemas A/B, P/F1 and P/F2 and appear in Figure 6.3 below.

Figure 6.3 Alternate Schemas and Factors used to Determine Borderline Grades

<table>
<thead>
<tr>
<th>Alternate Grading Schemas</th>
<th>Factors Used in Decision-Making</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Schema A/B</strong></td>
<td>Perceived Student Effort</td>
</tr>
<tr>
<td><em>A/B Border</em></td>
<td>Avoidance of Potential Negative Consequences</td>
</tr>
<tr>
<td><strong>Schema P/F1</strong></td>
<td>Understanding of Underlying Error and Inaccuracy in Classroom Grading</td>
</tr>
<tr>
<td><em>Pass/Fail: 1-2 Points</em></td>
<td>Perceived Student Effort</td>
</tr>
<tr>
<td></td>
<td>Avoidance of Potential Negative Consequences</td>
</tr>
<tr>
<td><strong>Schema P/F2</strong></td>
<td>Perceptions of Student’s Extenuating Circumstances</td>
</tr>
<tr>
<td><em>Pass/Fail: ~ 5 Points</em></td>
<td>Avoidance of Potential Negative Consequences</td>
</tr>
</tbody>
</table>
There were patterns in grading behavior among teachers in this study with respect to these factors or facets and their level of importance in decision making. This suggests that teachers framed different borderline grades differently. In the following sections, the alternate schemas for dealing with failing borderline grades, Schemas P/F1 and P/F2 will be discussed first highlighting the different types of reasoning used along the same grading border. Then differences in teachers’ grading practices and reasoning along the A/B border will be described by comparing the use of Schema A/B or a content-driven grading schema.

**Alternate Schemas of Grading for the Pass/Fail Border.** For most of their end-of-term grading, teachers relied on the numerical or proficiency grades they assigned students. However, when it came to the pass/fail border and sometimes along the A/B border, teachers did not use a content-focused schema to determine student grades. Along the pass/fail border, all of the participants in this study deferred to other factors to make their grading decisions. Teachers approached the pass/fail border in two different ways depending on how far the student’s end-of-term grade fell below the border.

**Failing Grades within 1-2 Points of Passing.** When students’ end-of-term grades were very close to passing, teachers used several specific non-achievement factors to determine the final grade. Teachers’ framing of this grading border and recognition of three specific non-achievement factors is referred to as alternate grading schema P/F1. All of the teachers recognized some level of inaccuracy in the grade that appeared in the grading software and questioned how accurate their grading was. Based on this inherent inaccuracy, teachers would bump up the grades of students slightly below passing. Rather than paying attention to the numerical grades (or the academic achievement of the student) only, teachers paid more attention to the small size of the margin by which students could receive a failing grade.
Additionally, teachers used two other pieces of reasoning in determining close pass/fail grades. They often brought up a second facet in their decision making: the potential negative impact or consequences for failing students. Additionally, teachers also considered how much effort the student put forth throughout the term. Broadly, when teachers decided to bump up failing grades, they used multiple facets of reasoning in a hierarchical level of importance. First, they considered the inaccuracy in their grading which was followed by their perception of a student’s academic effort. Finally, they sought to avoid perceived negative consequences for students.

This behavior is an example of schema switching when making decisions about borderline grades. For most borders, teachers used their classroom assignment grading system or rules and term calculation – a content-focused schema for evaluating students and determining grades. However, rather than continue to frame grading decisions based on content-focused gradebook calculations, teachers switched to this alternate frame (Schema P/F1) to decide on student grades on the pass/fail border. Teachers’ attention shifted from numerical calculation to the margin of potential error in their grades, student effort, and the potential negative consequences of a failing grade.

Failing Grades within 5 points of Passing. When students’ end-of-term grades were below the border for passing by more than 1 or 2 points, but less than 5 points, teachers considered the inherent inaccuracy in their grading, but no longer used it as a reasoning factor to bump up the final grade. They often felt that the failing grade was indeed failing, within their margin of error. This particular borderline case was more difficult for teachers to make decisions about. In some cases, teachers held the grade as is and failed the student. In other cases, however, teachers did raise the grade to passing. The most frequently cited reasoning facets in this case
were their understanding of the student’s extenuating circumstances and their desire to avoid potential negative consequences for the student. This frame for assigning grades utilizes two specific non-achievement factors and is referred to as alternate schema P/F2.

**Different Reasoning along the Same Border.** Teachers do not treat borderline grades along the pass/fail border in the same way. This behavior is indicative of two separate but consistent ways of framing these failing grades, rather than an idiosyncratic, haphazard approach. Teachers were very consistent in their practices along the 1–2-point margin of failing regardless of the subject taught or what type of grading policy their school used. When failing end-of-term scores fell outside that margin but were equal to or less than 5 points below the border, teachers made different choices. Some teachers held the grade as is, feeling that it was genuinely a failing score. When teachers made the choice to raise the grade to passing, there was a uniformity in how they rationalized that decision: their understanding of the extenuating circumstances a student might be experiencing and their perception of potential negative consequences for the student if they failed. Again, this reasoning process was explained similarly by the participants in this study.

**Alternate Schemas of Grading for the A/B Border.** Similar patterns of behavior exist along the A/B border, but there are also some interesting differences. Like failing borderline end-of-term grades, grades within 1 or 2 points of an A warranted some consideration. Approximately half of the teachers in this study adamantly insisted they do not raise end-of-term grades that are within 1 or 2 points of an A regardless of the situation. However, the remaining teachers said they do raise grades up to an A in that situation. Patterns of behavior and reasoning existed among both groups of participants (those that would hold the grade and those that would raise it).
Interestingly, neither of these lines of reasoning used error and inaccuracy-based reasoning along the A/B border despite using it along the pass/fail border.

**Holding B+ Grades.** As stated before, teachers would often consider the margin of error when determining the end-of-term grade for a potentially failing grade. Teachers that applied error and inaccuracy reasoning in deciding to bump up failing grades did not necessarily apply it to the A/B border. In fact, all of the teachers that did *not* round up on the A/B border *did* round up on the pass/fail border, using error and inaccuracy in their final grade as primary reason. The decision not to use this same reasoning along a different grading border is further illustrative that teachers may be framing “grading” in these cases differently, or through a different schema. The reasoning used to hold B+ grades indicated teachers were using a content-driven schema for grading rather than the alternative, Schema A/B. When teachers decided to raise B+ grades, they utilized two non-achievement factors.

**Raising B+ Grades.** For the teachers that did decide to bump students’ end-of-term grades up along the A/B border, the student’s effort was the driving consideration. Avoidance of perceived negative consequences for students often appeared as a secondary reasoning factor. The utilization of this reasoning is referred to as Schema A/B.

Many teachers would often look at students on the A/B border and determine the final grade based on how much effort the student put forth. Student effort was exhibited by the willingness to improve content understanding (in its own right), use of classroom opportunities to improve understanding and grades, the ability to ask for help and a focus on learning. Many teachers would often look through individual student grades and then make decisions on how they viewed the student’s effort throughout the course. These teachers reframed the determination of borderline A/B grades by focusing on student effort and avoidance of potential
negative consequences to determine the borderline grade rather than the numerical calculation created from a content-focused schema.

In summary, teachers who raised grades relied on their own understanding and acknowledgement of error and inaccuracy within their grading and also considered student effort. Interestingly, the teachers that did not raise grades did not use either of these reasoning factors. However, based on their descriptions of error, inaccuracy, bias, and subjectivity in their grading practices, these teachers presented it as a systemic issue. Hypothetically, these concerns existed for all grading borders. However, the error-based reasoning factor was far more prominent in decision making along the pass/fail border and remarkably absent along the A/B border. Half of the teachers in this study were willing to raise grades on the pass/fail border using error-based reasoning, yet they did not apply this same reasoning to the A/B border, regardless of whether they raise the grade or not. When deciding to raise grades at either border, teachers focused on students’ academic effort and tried to avoid negative consequences.

Teachers utilized non-achievement factors and alternate schemas (A/B, P/F1, and P/F2) in making decisions about the pass/fail and A/B border. In contrast, they approached other “middle” grade borders using their assessment of students’ content understanding throughout the term. Despite teachers’ consideration of certain factors in deciding on end-of-term grades (error/inaccuracy in grading, effort, minimization/avoidance of perceived negative consequence), these factors had varying levels of importance in different grading contexts.

**Viewing Borderline Grades as a Set of Contexts**

Based on the results from this study, it appears that “grading,” specifically borderline grading, is a collection of differently framed practices that teachers engage in. Some of teachers’ grading was focused on the evaluation of student content knowledge, or academic achievement,
which is the recommended practice. However, this content-focused schema was not utilized equally for all decision-making. B/C and C/D end-of-term grades were primarily determined based on students’ academic achievement; however this was less influential on the pass/fail and A/B borders.

Teachers used different facets of reasoning in differing orders of importance along different borders. Different factors were more or less relevant for different situations. For example, extenuating circumstances were a concern for students failing a course by a margin of 5 points rather than failing by 1 or 2 points or being 1 or 2 points below an A.

It appears that determining borderline end-of-term grades also fulfills functions other than communicating academic achievement. For students failing by 5 points, teachers tried to mitigate additional negative experiences for their students when possible. Fern summed it up nicely: “You don’t want to layer tragedies on tragedies.” Teachers used their understanding of students’ extenuating circumstances and the avoidance of negative consequences to make their decisions. In doing this, grading became a mechanism by which teachers could promote student well-being rather in addition to judging their academic performance. Students’ academic performance had already earned students a borderline end of term grade. When making the decision to bump up grades, perhaps teachers’ goals were to generally encourage rather than discourage students in regard to their schoolwork and attendance and to avoid more serious disengagement with their education.

When end-of-term grades were within 1 to 2 points of failing, the teachers in this study often questioned how accurate their grading was. Their focus was primarily on this inaccuracy which prompted them to consider additional factors: avoiding negative consequences for the student and students’ effort. Much like the previous example, grading along this border had
considerations beyond academic achievement. Teachers promoted student’s well-being when they chose not to fail a student and rewarded students who put in effort by passing them.

Similarly, along the A/B border, when teachers chose to bump up end-of-term grades, they focused on effort and the avoidance of negative consequences. Like the close failing margin, it appeared teachers both rewarded students and focused on their well-being in addition to their content knowledge.

To conclude, several factors have been identified that influenced teachers’ decisions about borderline grades. The time of year, specifically end-of-term grades were the only grades that teachers in this study considered to be true “borderline cases.” Within these borderline cases, the teachers in this study focused only on grades at the pass/fail and A/B border and consistently applied alternate schemas for grading. This is in contrast to prior research on grading practices where teachers pay attention to each grade or proficiency border. However, teachers’ focus on the pass/fail border and the practice of raising these borderline grades is quite consistent with prior research on grading factors.

The teachers in this study only raised end-of-term grades or held them, as opposed to lowering them. This practice contradicts some of the previous research, where teachers lowered grades in a small percentage of cases. However, in line with previous research, teachers primarily raise grades based on non-achievement factors, with a student’s academic effort being a large motivating factor. Additional influences included: perceptions of students’ extenuating circumstances, avoidance of negative consequences for the student, and concerns about error and inaccuracy in grading.

The emergent, overarching theme from this study is that teachers utilize different reasoning to make decisions about grading on different borders. This phenomenon might indicate
that teachers contextualize cases of end-of-term borderline grades differently and utilize different schemas or frames in determining the final grade. End-of-term borderline grading encompasses a variety of contexts for the teachers in this study which might explain larger trends in the literature of idiosyncratic and inconsistent grading. By utilizing a methodology of in-depth interviews with a narrower subset of teachers, focusing on teachers’ decision-making specifically for end-of-term borderline grades, and drawing on the concept of schema switching or frames, this study both builds on and extends prior research on K-12 teachers’ grading practices. In prior literature, several non-achievement factors have been identified and appear to be commonly used by K-12 teachers. However, this study shows that particular non-achievement factors have less or little to no influence on high school physical science teachers’ decisions about borderline grades. Additionally, other less explored concerns, like error in grading or students’ extenuating personal circumstances, are very influential to the decision-making process. Furthermore, despite the majority of prior research indicating that all borders receive similar attention, the results from the current study suggest the contrary. High and low borderline grades receive far more attention than grades in the middle of the grading range. The findings from this study support some aspects of previous research, but also identify areas for future research. The limitations of this study will be discussed in the next section. In later sections, implications for future work will also be discussed.

**Limitations of the Study**

Like any study, this one has limitations. The participant pool was small and consisted of teachers who were willing to be interviewed and recruited through emails and snowball sampling. Although the schools have been bracketed into three groups, these categories were
created after participant recruitment rather than recruiting participants from schools with particular grading policies.

Additionally, for each of the schools, I was only able to recruit a subset of the physical science teachers in the department. Therefore, teachers’ reports of their schools grading policies and their own decision-making might not be fully representative of the school or may be potentially skewed.

The majority of teachers in the current study taught upper level or advanced classes and electives. Student behavior may not have been a factor in their grading decisions because poor behavior occurs less frequently in these upper-level classes. Additionally, upper-level teachers possibly may have been more willing to participate in the study, which could potentially affect the study sample.

The interview for this study were conducted during a global pandemic, which might have also affected the results. Teachers were specifically asked about pre-pandemic experiences determining borderline grades; however, the teachers were actively working in pandemic. Concerns about the pandemic and their students could have affected how they described and self-reported their grading decisions.

The location of the study, within the context of a largely rural state, and a state that had recently pursued reform in assessment at the secondary level through a proficiency-based structure, may not reflect other state and school contexts, e.g., those that are more urban with much larger district administrations and teacher staffing in schools. Further research in other settings would be needed to investigate whether similar grading patterns exist.

However, despite the sampling constraints, the participants in this study exhibited common behaviors surrounding borderline grades. It is likely that the common practices and
experiences found in this study may be present in the larger population of secondary physical science teachers (and perhaps other disciplines). However, there could be a commonality shared among the teachers in this study that better predicts their similarities in practice. Despite working full-time as teachers during a global pandemic, teaching multiple modes of instruction and multiple classes to multiple students, these teachers agreed to meet and be interviewed twice after receiving my “cold call” recruitment email.

Teachers also may have been uncomfortable sharing certain kinds of practices or opinions related to grading students, especially with an unfamiliar person. This could be particularly important with respect to modifying borderline grades. Many teachers reported raising end-of-term student grades, but adamantly denied engaging in the practice of lowering end-of-term grades that were just above the border for a mark. It is possible that teachers may have been worried about portraying themselves in a critical light when describing decisions that led to lowering a student grade. However, as an experienced teacher conducting this research, I can confirm that the participants freely engaged in discussion with me with mutual collegiality. If teachers engaged in the practice of lowering grades, they may have felt comfortable enough to discuss it with a fellow teacher.

Interviewing as a methodology relies on people to recount their own experiences honestly and accurately in the presence of another person which could potentially influence what they discuss and how they describe it. The primary limitations in this study come from the small sample size, lack of generalizability in school type and self-reported nature of the data.

However, as described in Chapter 3, multiple steps were taken to increase the trustworthiness of this study and minimize bias. Feedback on the interviewing results, data, analysis, and interpretation from my doctoral committee was helpful throughout the research.
process. Additionally, my interpretations about grading practices were member checked with participants during interviews. My role as a practicing physical science teacher also increased the depth with which I was able to explore teacher experiences and contribute to research on grading practices in borderline situations.

**Implications and Recommendations**

In the following sections, implications from the study’s findings and suggestions for future research and changes to teacher practice and policy in Maine will be discussed. Although the data from this study are limited to the context of Maine and only secondary physical science teachers, the findings indicate several areas for future research, teacher practice, and school and state approaches to policies on grading.

**Implications for Future Research**

Three approaches to school grading policy were described in this study: proficiency-based grading, a school with many district grading guidelines, and schools with fewer district guidelines. These categories were developed throughout the research process to address the research questions. An interesting area of future research would be to investigate the grading practices of teachers from known school structures and grading contexts with the purpose of qualitatively exploring differences or similarities in teacher practice. This would help better refine our understanding of the grading context and policies teachers work within and better establish if the patterns observed in teachers’ decisions-making in this small study are representative of larger grading trends in Maine and elsewhere.

Additionally, this study focused on teacher practices and general decision-making about borderline end-of-term grades. Teachers were not asked about the value systems they considering in determining grades or the values they might try to convey with grades. Thus, this
study did not attempt to understand each teacher’s philosophical views about grading generally, or grading in the discipline of science. Teachers preferentially focused on the ends of the grading spectrum (pass/fail and A/B) while not considering student grades within the middle: C/D and B/C. It is possible that these “middle of the road” students may not be receiving equal levels of attention, which could potentially affect classroom equity. This study identified borders that were important to teachers in their grading, but it would be worthwhile to study why certain borders were more or less important than others, specifically with respect to the values teachers assigned to them.

All of the schools in this study utilized online grading platforms to communicate with students, parents, and stakeholders about student grades. Rather than always post grades as they appeared on grading portals, teachers engaged in the process of raising end-of-term grades despite the apparent transparency of communication. This process of behavior might speak to larger societal and community expectations about how online grades are perceived and treated. The exclusive use of online grading practices is a relatively new practice. (For example, when I started my own teaching career in 2008, my district was transitioning away from the use of private, paper-based grade keeping and towards the use of online grading platforms.) Although these grading platforms aid in the communication of grades and student performance, they may have some potential, unintended consequences for teachers’ grading practices. Teachers raised concerns both about the mechanics and constraints imposed by computerized grading systems, as well as the public oversight of their grading records, and indicated these factors do influence their grading behaviors and practices. These are potentially rich topics for future research.
Implications for Teacher Practice and Professional Development

By interpreting grading as a set of differently framed activities, rather than a single activity, teachers’ grading practices can be interpreted with a more nuanced view. Teachers can struggle to change their grading practices despite having professional development (Olsen and Buchanan, 2019). This study and prior research illustrate the limited impact training may have as teachers’ continue to exercise individual judgement in how to and to what extent follow district or school guidance on grading in specific cases. Thus, grading practices continue to be quite variable by teacher, with a few common patterns. Additionally, the schools in this study employed three main types of approaches to grading which had only a limited effect on teachers’ practices around borderline grades. Ideally, per experts, grading should solely be a measurement of academic achievement. However, in practice, the process of borderline grading utilizes multiple frames or schemas rather than only a focus on academic achievement. It is possible that because grading is presented exclusively as the evaluation of academic achievement rather than the multiple forms it takes, changes in practices are slow or do not occur at all. One suggestion for future research would be to further investigate the origins and functions of the different schemas teachers utilize in their grading practices. If this is better understood perhaps professional development and grading reforms can be modified to better address the complicated practice of classroom and end-of-term grading.

Implications for Policy

The vast majority of schools in this study used numerical grading with fewer district guidelines which could be representative of districts or schools in Maine as a whole or not. Despite reform efforts and legislation to award proficiency-based diplomas, many of the schools in this study did not determine end-of-term grades using the proficiency-based approach.
Furthermore, the teachers in this study exhibited similar decision-making in determining borderline end-of-term grades despite of the different grading policies of their schools.

Numerical scales, even with specifications about the weights of relative grading, seemed to promote the use of non-achievement factors in the end-of-term grading decisions. A students’ end-of-term score simply became a number – a number that teachers recognized as having some range of error inherent within it and was constrained by what they felt were arbitrary boundaries.

Ultimately, Seacoast High School, which shifted completely to the proficiency-based grading structure with Maine’s LD.1422 legislation, struggled with these issues as well. In practice, their proficiency scale became similar to a 4-point GPA system with student proficiencies across standards being averaged over the course to produce a single number that represented a student’s academic achievement. However, proficiency-based grading may better promote grading for academic achievement and increase grading equity. Despite engaging in patterns of behavior similar to that of all the other participants, Piper and Paul from Seacoast seemed to focus more on students’ learning and ability to demonstrate proficiency on the NGSS Science and Engineering practices up until they needed to post the final end-of-term grade on their online grading software.

The distillation of student learning to a single value, regardless of how it is calculated, undermines the complexity of learning and promotes teachers’ use of non-academic factors in this final end-of-term decision making. Teaching and learning is a complex process and grading should perhaps better reflect that process by not minimizing learning over the course of a term to a single value.
Implications for Equity

The teachers in this study utilized additional factors, not only academic achievement, to determine students’ end-of-term borderline grades. These factors were applied differently to different borders, which suggests that teachers apply different schemas of reasoning to borders and use different perceptions of students’ attributes or situations. Some of these choices were made from compassion for students. Paul described his choice to bump up borderline grades: “I would call it compassionate, but somebody might say, lenient or something like that.” Teachers recognized inequities in students’ situations and less than ideal circumstances for learning. Thus, they tried to minimize potential negative consequences or reward certain positive student attributes (i.e., effort).

Teachers’ grading and decision-making is culturally situated. Unfortunately, despite teachers’ efforts to increase their perceived equity in grading, teachers might unconsciously be contributing to inequity in grading. Several decision-making factors (extenuating circumstances and effort) and the preferential attention paid to certain borders have important implications for equity.

Through conversations with students, parents, guidance counselors, or other stakeholders, teachers were aware of students’ extenuating circumstances in particular instances. In all likelihood, there were student situations that teachers were unaware of. When teachers considered students’ extenuating circumstances, they did not apply this equally to all students due to this lack of information. Not knowing about students’ extenuating circumstances could be the result of several reasons. A lack of information could be due to quiet students not sharing challenges, health issues, or non-ideal situations for learning. Sharing details about family life and or outside influences that had negative educational impacts could also be cultural in nature –
with certain students or stakeholders finding it inappropriate to share certain information, while others did.

Teachers’ perceptions and consideration of student effort also bring up potential issues in grading equity. Effort is perceived as a “good” student quality. Effort and other favorable student qualities are culturally situated. It would be worthwhile to examine these attributes and how they contribute to inequities in grading. Additionally, students that can exhibit effort in their coursework might have a certain amount of time to be able to do that or come from certain home lives. Effort can be perceived in a variety of ways, one of which is homework completion. There is the general perception in teaching that time spent on homework varies by socioeconomic (SES) level, with students at lower SES levels having less time to work on homework. However, some studies argue the contrary, stating time spent on homework is not affected by students’ and parents’ socioeconomic level; rather, more nuanced qualities, like homework efficacy are (Daw, 2012). The concept of student “effort” is multidimensional and warrants further exploration, specifically with respect to equity.

Lastly, the lack of attention paid to the B/C and C/D borderline grades implies that this particular group of students is often overlooked and not given the same consideration as their peers. B+ students will sometimes get their end of term grades bumped up, as will students that are close to passing. B/C and C/D students are not afforded this chance. It is possible that this grading practice might serve as a demotivator for this particular group of students and potentially increase grading inequities because more attention is paid to grading borders and demographics of students.
Conclusions

In conclusion, this chapter examined the findings of the research with respect to prior literature on teachers’ grading practices. Teachers consistently used their perception of students’ academic effort to make decisions about whether to raise students end-of-term borderline grades or hold those grades as is. This is consistent with prior literature which shows that student effort is a frequently used non-achievement factor that teachers use to determine grades. Additionally, teachers often either decided to raise grades or hold them as is. Raising and lowering of grades have been found in prior research, however, raising or holding of grades was the more frequent teacher practice in this study. In contrast, none of the teachers in this study described lowering grades and were adamantly against this practice, contradicting previous findings.

Teacher perceptions about student behavior and ability were not highly influential on teacher practices despite having documented effects in prior research spanning K-12 grade levels. It is possible that high school teachers engage in grading practices that are different from K-8 practices because high school teaching and grading occur in a much different context. The lack of attention to these two non-achievement factors could also be a result of decreased incidents of “poor” student behavior due to truancy or poor attendance. Additionally, high school classes are typically tracked and multiple subject areas have several different levels of instructional rigor (e.g., regular, honors, Advanced Placement). Ability may not factor into the grading decisions of high school teachers because students are tracked into classes with other students of similar ability levels.

The teachers in the current study frequently mentioned concerns about error or inaccuracy in their grading. Concerns around error and inaccuracy in grading have been frequent topics of study in prior grading and educational measurement research. However, teachers’ views
about inaccuracy and error remains and unstudied area that could be a focus of future research. Additionally, it might be possible that as science teachers, the participants in this study utilized their understanding of measurement error to help determine end-of-term grades. Teachers of other disciplines may not have similar concerns or self-reported experiences. However, this is an area of uncertainty in this study and a very interesting area of further research.

The different grading guidelines or scales used by the schools in the current study appeared to have little to no effect on teachers’ borderline grading decisions or the factors that influenced them. Teachers used a variety of factors in determining borderline grades, consistent with previous literature. There are challenges in distilling the assessment of students’ content knowledge down to a single numerical value; perhaps teachers’ incorporation of non-achievement factors, regardless of the grading system used, are indicative of that.

Lastly, grading remains a challenging area of teacher practice and educational research. Despite recommendations for grading reforms or best practices, grading, as it occurs in the American school system is very slow to change. Grading often is described as fulfilling a single function in the literature: assessing students’ academic achievement. However, in practice, grading takes on a variety of roles in teachers’ classrooms, in the schools and districts they work in, and in the lives of their students. Borderline grading decisions make these different aspects of grading highly visible and suggest teachers apply multiple frames when making decisions.

Teacher decision-making about borderline grades was described as different framing situations, which caused teachers to utilize different reasoning and apply different schemas to different borderline grading scenarios. Grading for teachers, especially the practice of determining borderline grades, appears to be a more complex and varied set of grading contexts, rather than a generic exercise – with teachers using different reasoning elements differently in various
situations. Through a qualitative, interview-based approach to understanding teachers’ self-reported experiences, a more nuanced understanding of teachers’ grading practices has emerged from this study. These results help to explain some of the well-documented variation in grading practices in a more productive and teacher-positive light.
REFERENCES


APPENDIX A

Initial Recruitment Email

Dear Teachers,

This email is to request your participation in a research project I obtained your email from a colleague (include name of the colleague) who suggested you might be a good candidate for this particular project. In addition to being a full time teacher, I am currently a doctoral student in the College of Education and Human Development at the University of Maine, Orono. As part of my dissertation research, I am interested in interviewing high school physical science teachers with three years of prior experience about their assessment and grading experiences within their school setting. The attached consent form includes more information about this study.

This study will consist of two virtual interviews to take place during the 2020-2021 school year. The first interview will take place during the winter of 2021 and the second will be approximately two to three months after. These interviews will last approximately one hour each and be recorded via Zoom. These will be transcribed, and I will assign you and your school pseudonyms to be referred to throughout the research. Any identifiable information in the transcript will be removed. All data during video-recorded interviews and all transcripts will be kept on password protected, encrypted hard drives. Audio and transcript data will also be archived on an online password protected drive that meets FERPA standards. All video and audio data will be destroyed when the study is completed. The expected date of study completion is August 2022. Transcripts may be kept indefinitely for reference.

Your participation in this study is voluntary. There are no foreseeable risks from this study. You can choose not to answer any interview questions at any time. You can withdraw from the study at any time. If at any time during the interview or the study, you decide to withdraw from the study, all recordings and transcripts associated with you will be destroyed.

If you have any additional questions about this study, please contact me at (email address) or my advisor, Michael C. Wittmann (email address).

If you are willing to be interviewed, please contact me for scheduling. If you agree to be interviewed, I will send you a reminder email two days prior to each interview. Thank you for your time and consideration.

Sincerely,

Kate Hayes
APPENDIX B

Recruitment Email for “Snowball” Participants

Dear Teachers,

This email is to request your participation in a research project I obtained your email from a colleague (include name of the colleague) who suggested you might be a good candidate for this particular project. In addition to being a full time teacher, I am currently a doctoral student in the College of Education and Human Development at the University of Maine, Orono. As part of my dissertation research, I am interested in interviewing high school physical science teachers with three years of prior experience about their assessment and grading experiences within their school setting. The attached consent form includes more information about this study.

This study will consist of two virtual interviews to take place during the 2020-2021 school year. The first interview will take place during the winter of 2021 and the second will be approximately two to three months after. These interviews will last approximately one hour each and be recorded via Zoom. These will be transcribed, and I will assign you and your school pseudonyms to be referred to throughout the research. Any identifiable information in the transcript will be removed. All data during video-recorded interviews and all transcripts will be kept on password protected, encrypted hard drives. Audio and transcript data will also be archived on an online password protected drive that meets FERPA standards. All video and audio data will be destroyed when the study is completed. The expected date of study completion is August 2022. Transcripts may be kept indefinitely for reference.

Your participation in this study is voluntary. There are no foreseeable risks from this study. You can choose not to answer any interview questions at any time. You can withdraw from the study at any time. If at any time during the interview or the study, you decide to withdraw from the study, all recordings and transcripts associated with you will be destroyed.

If you have any additional questions about this study, please contact me at (email address) or my advisor, Michael C. Wittmann (email address).

If you are willing to be interviewed, please contact me for scheduling. If you agree to be interviewed, I will send you a reminder email two days prior to each interview. Thank you for your time and consideration.

Sincerely,

Kate Hayes
APPENDIX C

First Interview Protocol

Thank you for agreeing to participate in my dissertation research. I’d like to remind you that this is being recorded. Over the next hour, I’m going to ask you some questions about how you experience and make decisions about grading and assessment in your classroom and your school. You are free to stop the interview at any time or decline to answer any question I ask. Do you have any questions before we start? I’d like to start off with some general background questions.

1. How long have you been at your current school?
2. Do you have a degree in this (science) content area? Or a teaching degree?
3. How long have you been teaching in this content area?

As I mentioned in the recruitment email, my research is on teacher experiences of grading and assessment within their classrooms and schools. I’d like to first ask you what you think about grading and assessment.

4. Do you consider assessment and grading the same thing? Why or why not? (follow-ups for clarification as needed)

Now I’m going to ask you assessment and grading practices in your classroom before we experienced the COVID-19 pandemic since a lot has changed in the past year. I’ll ask you separately about assessment and grading for most questions, but I understand if there is overlap between those. If you feel like I’m asking you the same question, please let me know. I’m especially interested in how these aspects of teaching overlap for you so please elaborate as you need to. I’d like to start by asking you about assignments and student work…

5. What kinds of assignments or student work do you assess in your classroom? Why?
6. What kinds of assignments or student work do you grade in your classroom? Why?
7. What kinds of student work do you not grade or assess? Why?
8. Ideally, what would you assess? And what would you ideally grade?
   a. What prevents you from that ideal?
   b. What would need to be different to help you get closer to your ideal?

You touched on some factors that influence your grading and assessment practices in the last question. The next set of questions focuses specifically on external factors in your school that influence these practices. Again, I would like you to focus on your experiences before the COVID-19 pandemic.

9. What external factors influence your assessment practices –such as expectations at the school or district level? Please explain.
10. What external factors influence your grading practices? Please explain.
11. How do your own views and beliefs about teaching and learning influence your assessment and grading practices?
12. Can you describe a time when you questioned your own decisions about assessment and/or grading? Explain.
   a. Follow-up question about assessment or grading if participant only focused on one.
13. Can you describe a time when your assessment and/or grading decisions were questioned or challenged by others in your school or district? Explain.
   a. Follow-up question about assessment or grading if participant only focused on one or the other.
   b. How was this resolved?
   c. How did you feel about the resolution of the situation? Would you have preferred a different outcome? Why?

We’re going to switch topics now and discuss your experience of grading and assessment during the pandemic, specifically as many schools transitioned to remote learning.

14. How did your grading and assessment of student work change as your school transitioned to remote learning?
15. What kinds of assignments or student work do you assess in your remote classroom?
16. What kinds of assignments or student work do you grade in your remote classroom?
17. What kinds of remote student work do you not grade? Why?
18. How did your grading and assessment of student work change as your school transitioned to remote learning?
19. What factors influenced your choices of what to grade and what not to grade?
   a. What challenges concerning assessment and grading did you face during the spring of last year?

I have a few more questions and then we’ll be done with the interview.

20. What would your ideal remote grading system look like? How about remote assessment?
21. What constraints prevent this ideal?

We’re almost done. Thank you so much for agreeing to be interviewed. I want to make sure I addressed everything you wanted to talk about.

22. Is there anything you would like to add that I did not ask you about already?

Alright. Thank you so much for your time. I will be in contact with you in the next several weeks to schedule a second interview.
APPENDIX D

Second Interview Protocol

The last time we spoke, I asked you about your grading and assessment practices as a science teacher. Today I’d like to focus specifically on grading and your experience giving numeric scores or grades to students’ work, and what influences your decision-making, in borderline cases. By borderline, I mean when a score falls just above or below the border for a mark or proficiency level. In answering the questions, I’d like you to think about the range of different kinds of student work that you score (homework, labs, projects, quizzes, tests, and report cards/course grades).

To start off, let’s address some background information I’m not sure if we discussed last time.

1.) What type of grading system does your school use? What is the cut off or border between marks?

2.) Please describe an experience you remember, or a situation, where you scored a student's work and it fell just below the border for a mark (i.e. – A, B, C, etc). What does this experience feel like for you? Did you (or do you) deliberate over the final score to give the student? Please explain why. What do you usually do in this type of case?

3.) What different factors do you consider when making the final decision about the score/grade in a case like this?

4.) In a case where a student's numeric score is just below a mark or proficiency level, do you generally raise the score to the next level? Please explain your decision-making process for this type of situation.

5.) In a case where a student's numeric score is just above a mark, letter grade, or proficiency level, do you let that score stand or adjust it? Follow up: Please explain your decision-making process for this type of situation.

6.) Are there certain types of student work that you find more challenging to assign final scores when they are just under the border for a mark, letter grade, or a proficiency level? Please explain what these are.

7.) How do you decide on a final score/grade for high stakes assignments that are just under a score for a proficiency level? Follow up: What factors influence your decision making in this type of situation?

8.) Have you ever experienced any of the conflicts or factors you’ve mentioned above in a non-borderline situation? Follow up: What did you decide to do and why?
9.) In the previous interview, you had mentioned some beliefs that influenced your assessment and grading practices. How, if at all, do these beliefs influence your grading or assigning of grades to borderline situations?

10.) In the previous interview, you had mentioned some factors factor that affected your grading practices. How, if at all, do these factors influence how you grade borderline situations?
BIOGRAPHY OF THE AUTHOR

Kate Roberts attended Northeastern University in Boston, MA. While there, she received several awards: the 2004, 2005, and 2006 Joseph Arthur Coolidge Achievement Award; the 2003, 2004, and 2006 Lawrence Award for Scholastic Excellence in Undergraduate Physics; the 2006 President’s Award; and the 2006 Condit Award.

She graduated *summa cum laude* with a Bachelor of Science in Biomedical Physics degree in May of 2006. She was Class Marshall (valedictorian) for the College of Arts and Sciences graduating class that year.

Kate returned to Maine in the summer of 2006 to become a high school teacher. She became a Knowles Science Teaching Fellow in 2007 and, after graduating from the program, a senior fellow in 2012. She earned her Master of Science in Teaching from the University of Maine in 2009 under Michael Wittmann. Her thesis was titled: “A Qualitative Analysis of Student Behavior and Language During Group Problem Solving.”

Kate has been at high school teacher since 2008 and works at Bangor High School in Bangor, Maine. Kate has been a finalist for the Penobscot County Teacher of the Year twice. She was also recognized by the University of Chicago as an “Inspiring Educator” in 2021. Kate enjoys teaching, but often questions her choice of career. She possesses an abundance of frustration and feels teachers need to be a more prominent factor in educational decision-making. Kate loves her students, but often finds herself without the time, resources, or support to prioritize their learning and high school experience. She deeply values her doctoral work as a way elevate the voices of teachers and hopefully provide insight into a narrow, but complex area of teacher practice. She candidate for the Doctor of Philosophy in Education with a concentration in STEM Education from the University of Maine in December 2022.