

A CRITICAL PEDAGOGY PERSPECTIVE OF THE IMPACT OF
SCHOOL POVERTY LEVEL ON THE TEACHER GRADING
DECISION-MAKING PROCESS

by

Joshua P Kunnath

B.A. (Michigan Technological University) 2004

M.A. (Oakland University) 2009

M.A. (California State University, Bakersfield) 2011

A dissertation

submitted in partial fulfillment of the requirements for the degree of

Doctorate in Education

Doctoral Program in Educational Leadership at Fresno State

Kremen School of Education and Human Development

California State University, Fresno

2016

Joshua P Kunnath
January 2016
Educational Leadership

A CRITICAL PEDAGOGY PERSPECTIVE OF THE IMPACT OF
SCHOOL POVERTY LEVEL ON THE TEACHER GRADING
DECISION-MAKING PROCESS

Abstract

This dissertation examined the effects of school poverty level on the teacher grading decision-making process. Utilizing two theoretical frameworks—a critical pedagogy framework primarily based on the work of Paulo Freire (1996) and a teacher grading decision-making framework from the work of McMillan (2003)—the study sought to compare teacher grading practices, the influences of teacher grading, and teacher grading rationale by low-, mid-, and high-poverty schools in order to determine poverty’s impact on the process. A small, yet significant number of studies have found grades in high-poverty schools to be less accurate and more subjective than low-poverty schools in communicating student achievement. This study contributes to the literature by helping to establish the link between high-poverty schools and inaccurate student grades. Additionally, this study is novel in being the first (to the knowledge of the author) to combine the concepts of teacher grading, poverty, and decision making.

Set in an ethnically and economically diverse high school district in California’s South San Joaquin Valley, this study employed a transformative explanatory sequential mixed methods design to answer a guiding research question: How does school poverty level affect the teacher grading decision-making process? A 36-item grading survey was used to collect quantitative data

from 251 teachers and additional qualitative data from 121 of these teachers. Focus groups were conducted at four schools sites to collect qualitative data from a total of 15 teachers. Quantitative data were analyzed with one-way ANOVAs, a MANOVA, Kruskal-Wallis tests, and descriptive statistics, while qualitative data were analyzed with a constant comparative analysis method.

Results indicated that the direct effects of school-poverty level on the teacher grading decision-making process are intricate and nuanced, yet existent and influential. The findings explained that teachers seek to grade in objective, pedagogically-sound ways that align to their own philosophy, but influences—both internal and external to the classroom—cause them to stray from these practices. These influences occur more often in high-poverty schools, and this leads to greater subjectivity and less accuracy of student grades. Models for interpreting teacher grading rationale and the impact of poverty on teacher grading decision making are presented.

Copyright by
Joshua P Kunnath
2016

California State University, Fresno
Kremen School of Education and Human Development
Doctoral Program in Educational Leadership

This dissertation was presented

by

Joshua P Kunnath

It was defended on

December 18, 2015

and approved by:

Mahmoud Suleiman, Chair
Teacher Education, CSU Bakersfield

Juan Carlos Gonzalez
Educational Leadership, CSU Fresno

Blanca Cavazos
Superintendent, Taft Unified High School District

DEDICATION

*To Dr. Insko, who brought me to Emerson,
and to Go Eun, who brought Emerson to me.*

ACKNOWLEDGMENTS

First and foremost, I would like to thank those without whom I would be neither in this world nor in a position to produce this dissertation: my family, both nuclear and extended.

I would like to extend sincere thanks and appreciation to my dissertation chair, Dr. Mahmoud Suleiman, who provided a great balance of guidance and freedom on this dissertation journey that was crucial to its completion and any possible success. Additionally, I would like to thank my committee members, Dr. Juan Carlos Gonzalez and Dr. Blanca Cavazos, for their recommendations and support throughout the process. Thank you to Dr. Dean McGee for your insight and words of advice along the way and to Dr. Danny Whetton for your support in adverse times.

Thanks go to my school site administrators, who provided support and assistance—especially in the later, challenging phases. I extend additional appreciation to the participants in the study, the cooperating administrators, and to the teachers at my own school site for all of your time and input.

Finally, I would like to thank my cohort, B³, for everything along this journey, but most of all for their friendship.

TABLE OF CONTENTS

| | Page |
|---|------|
| LIST OF TABLES | xii |
| LIST OF FIGURES | xiv |
| CHAPTER 1: INTRODUCTION | 1 |
| Statement of the Problem..... | 2 |
| Background | 3 |
| Theoretical Perspective | 6 |
| Critical Pedagogy Framework..... | 7 |
| Teacher Grading Decision-Making Framework | 12 |
| Significance of the Study | 18 |
| Purpose of the Study | 19 |
| Research Questions..... | 20 |
| Summary | 20 |
| Definition of Key Terms..... | 21 |
| CHAPTER 2: REVIEW OF THE LITERATURE | 24 |
| The Impact of Poverty on Student Achievement..... | 24 |
| Understanding the Income Achievement Gap | 25 |
| Explaining the Income Achievement Gap | 29 |
| Grading in American Public Schools..... | 50 |
| A History of Grading Practices and Concerns | 50 |
| Grading Today..... | 56 |
| Teacher Grading Decision Making..... | 73 |
| Teacher Grading Decision-Making Process..... | 74 |

| | |
|--|-----|
| Teacher Preparation and Training | 75 |
| Teacher Feelings and Roles..... | 77 |
| External Pressure | 79 |
| Classroom Context | 80 |
| The Impact of Poverty on Teacher Grading | 81 |
| Differences across School Poverty Levels | 81 |
| Differential Grading | 83 |
| Predictions | 84 |
| Summary | 85 |
| CHAPTER 3: METHODOLOGY..... | 87 |
| Purpose of the Study | 87 |
| Research Questions..... | 88 |
| Research Design..... | 89 |
| Quantitative Data..... | 93 |
| Qualitative Data..... | 94 |
| Participants/Sample..... | 95 |
| Sampling Rationale | 97 |
| Data Collection and Instrumentation | 98 |
| Survey..... | 99 |
| Focus Groups..... | 103 |
| Pilot Study..... | 104 |
| Basis of Comparison | 106 |
| Data Analysis | 108 |
| Quantitative Data..... | 108 |
| Qualitative Data..... | 109 |

| | |
|--|-----|
| Data Mixing..... | 110 |
| Limitations | 110 |
| CHAPTER 4: RESULTS | 112 |
| Introduction..... | 112 |
| Demographics of the Sample | 114 |
| Quantitative Data Analysis | 115 |
| Grading Practices | 119 |
| Influences of Teacher Grading | 129 |
| Qualitative Data Analysis | 131 |
| Promotion of Student Success..... | 136 |
| Grading Philosophy and Beliefs..... | 147 |
| Influences of Grading Practices | 154 |
| Grading Practices | 163 |
| Summary of Findings..... | 172 |
| CHAPTER 5: SUMMARY/DISCUSSION/CONCLUSION | 174 |
| Summary and Discussion of Findings | 175 |
| Research Question 1 | 176 |
| Research Question 2..... | 181 |
| Research Question 3..... | 187 |
| Guiding Research Question..... | 193 |
| Implications for Policy and Practice | 196 |
| Recommendations for Future Research..... | 199 |
| Limitations of the Study..... | 200 |
| Conclusion | 201 |
| REFERENCES | 203 |

| | |
|---|-----|
| APPENDIX A: SURVEY INSTRUMENT | 223 |
| APPENDIX B: INTERVIEW PROTOCOL..... | 239 |
| APPENDIX C: OBSERVATIONAL PROTOCOL..... | 241 |

LIST OF TABLES

| | Page |
|---|------|
| Table 1 <i>Teacher Grading Decision-Making Framework by Domain</i> | 14 |
| Table 2 <i>Alignment of Research Questions with Theoretical Frameworks</i> | 89 |
| Table 3 <i>Alignment of Research Questions, Methods, and Instruments</i> | 91 |
| Table 4 <i>A Comparison of High School Population Size, Diversity, and Poverty Rate by Level of Organization</i> | 97 |
| Table 5 <i>Alignment of Research Questions, Teacher Grading Decision-Making Framework, Instruments, and Analysis</i> | 100 |
| Table 6 <i>Alignment of Focus Group Questions by Research Question and Teacher Grading Decision-Making Framework Domain</i> | 105 |
| Table 7 <i>School Poverty Level Groups</i> | 107 |
| Table 8 <i>Demographic Information by School</i> | 116 |
| Table 9 <i>Survey Responses by School</i> | 117 |
| Table 10 <i>Survey Responses by School Poverty Level</i> | 118 |
| Table 11 <i>Demographics of Survey Participants by School Poverty Level</i> | 119 |
| Table 12 <i>Means, Standard Deviations, and Medians Comparing Teacher Assigned Grades by School Poverty Level</i> | 120 |
| Table 13 <i>ANOVA for Teacher Assigned Grades by School Poverty Level</i> | 121 |
| Table 14 <i>Means, Standard Deviations, and Medians Comparing Teacher Grading Practices by School Poverty Group</i> | 122 |
| Table 15 <i>Kruskal-Wallis Test Results of Grading Practices by School Poverty Group</i> | 125 |
| Table 16 <i>Mann-Whitney U Test for Teacher Grading Practice 14: Quality of Student Completed Homework</i> | 126 |
| Table 17 <i>Mann-Whitney U Test for Teacher Grading Practice 16: Student Extra Credit for Academic Performance</i> | 127 |
| Table 18 <i>Cronbach's Alpha for Teacher Grading Practices</i> | 128 |

| | | |
|----------|---|-----|
| Table 19 | <i>Means, Standard Deviations, and Medians Comparing Subjective Grading Index by School Poverty Level</i> | 128 |
| Table 20 | <i>ANOVA Results for Subjective Grading Index by School Poverty Group</i> | 128 |
| Table 21 | <i>Means, Standard Deviations, and Medians Comparing Influences of Grading by School Poverty Group</i> | 130 |
| Table 22 | <i>Multivariate Tests for Influences of Grading by School Poverty Level</i> | 132 |
| Table 23 | <i>Tests of Between-Subjects Effects for Influences of Grading by School Poverty Level</i> | 133 |
| Table 24 | <i>Description of Focus Group Teacher Make-up</i> | 134 |
| Table 25 | <i>Description of Qualitative Survey Participation</i> | 135 |
| Table 26 | <i>Emergent Themes and Frequency of Data References</i> | 136 |
| Table 27 | <i>Sub-Themes and Categories of the Theme Promotion of Student Success by School Poverty Level</i> | 137 |
| Table 28 | <i>Sub-Themes and Categories of the Theme Grading Philosophy and Beliefs by School Poverty Level</i> | 148 |
| Table 29 | <i>Sub-Themes and Categories of the Theme Influences of Grading Practices by School Poverty Level</i> | 154 |
| Table 30 | <i>Sub-Themes and Categories of the Theme Grading Practices by School Poverty Level</i> | 163 |
| Table 31 | <i>Summary of Findings for Research Question 1</i> | 177 |
| Table 32 | <i>Summary of Findings for Research Question 2</i> | 182 |
| Table 33 | <i>Summary of Findings for Research Question 3</i> | 188 |

LIST OF FIGURES

| | Page |
|---|------|
| <i>Figure 1.</i> Teacher grading decision-making framework..... | 13 |
| <i>Figure 2.</i> Mixed methods study design..... | 93 |
| <i>Figure 3.</i> Three significant components of teacher grading rationale. | 189 |
| <i>Figure 4.</i> The impact of school poverty level on teacher grading decision making..... | 195 |

CHAPTER 1: INTRODUCTION

Grades are the primary tool that schools use to communicate student success. While the general assumption is that they universally represent academic achievement, is this really the case? Prominent grading reformers such as Thomas Guskey, Robert Marzano, Doug Reeves, and Rick Stiggins insist that current teacher grading practices are a significant problem in schools (Guskey, 2015; Marzano, 2000; Reeves, 2011; Stiggins, Frisbie, & Griswold, 1989). In the first page of *Transforming Classroom Grading*, Marzano (2000) boldly declared, “grades are so imprecise that they are almost meaningless” (p. 1). Researchers in the field are in consensus that grades often fail to communicate achievement and may instead represent a mix of factors that varies depending upon the teacher (Bowers, 2011; Cizek, Fitzgerald, & Ranchor, 1995/1996; Guskey, 2015; Marzano, 2000; McMillan, 2001; Reeves, 2011; Stiggins et al., 1989). Not to be ignored in this discussion is the academic achievement gap that exists between children of families of high and low income (Borg, Borg, & Stranahan, 2012; Guskey, 2011; Heckman, 2006; Reardon, 2011; Sirin, 2005). Researchers have studied many factors that explain this gap, and a relatively small, yet significant body of research has pointed to teacher grading practices as a possible contributing factor (Agnew, 1985; Cauley & McMillan, 2000; Howley, Kusimo, & Parrott, 2000; McMillan, Myran, & Workman, 2002; Office of Educational Research and Improvement, 1994; Zwick & Himmelfarb, 2011). Although studies connecting teacher grading practices and the income achievement gap are far fewer in number than those explaining other factors contributing to the gap, possible implications of a link between the income achievement gap and ineffective grading practices are too significant to dismiss.

Statement of the Problem

By and large, teacher grading practices are highly subjective and differ from researcher recommendations. Instead of exclusively using student academic achievement to create student grades as recommended by researchers, teachers often use varying combinations of academic achievement and a handful of non-academic factors such as effort, behavior, and attitude. These practices result in grades that fail to attain the purpose for which they were created, which is primarily to communicate student academic achievement (Allen, 2005; Brookhart, 1991, 1994, 1999, 2009; Cizek et al., 1995/1996; Cross & Frary, 1999; Frary, Cross, & Weber, 1993; Guskey, 2015; Marzano, 2000; McMillan, 2001; Randall & Engelhard, 2010; Reeves, 2011; Stiggins et al., 1989; Wormeli, 2006). When teachers use non-academic factors to create student grades, the result can be either grade inflation or deflation, often depending upon the poverty level of the school (Kelly, 2008; Howley et al., 2000; Office of Educational Research and Improvement, 1994). However regardless of student success on these non-academic factors, the resulting grades may inaccurately communicate academic achievement. As a result, low achieving students may not be accurately identified, and unidentified students may not be provided opportunities to learn deficient skills (McMillan, 2001). These poor grading practices are likely more common in high-poverty schools, as teacher grading practices seem to face greater influence from internal and external pressures in high-poverty schools than in low-poverty schools (Agnew, 1985; Cauley & McMillan, 2000; Howley et al., 2000; McMillan et al., 2002; Office of Educational Research and Improvement, 1994; Zwick & Himmelfarb, 2011). This is not only a matter of poor practice, but also a matter of equity (Office of Educational Research and Improvement, 1994).

Background

Despite many apparent issues with current grading practices, grading reform is an extremely controversial subject in high school education. This controversy originates from a number of major sources: (a) grading constitutes one of the oldest education traditions in the United States (Brookhart, 2009; Kirschenbaum, Simon, & Napier, 1971; Smallwood, 1935); (b) current grading practices throughout the country are largely accepted by educators, students, parents, and the general public (Marzano, 2000; Office of Educational Research and Improvement, 1992; Tyack & Tobin, 1994); and (c) student grades influence so many aspects of the education system (and beyond) that significant reform likely would result in changes to the entire American education system with possible implications to society as a whole (Guskey, 2015; Marzano, 2000; Tyack & Tobin, 1994). Regardless of the abundance of research stating that student grades should be based only on student academic achievement, both educators and the entire school community are often hesitant and resistant to change current grading practices (Dueck, 2014; Erickson, 2010; Marzano, 2000). Instead, a large amount of education reform over the past 15 years has focused on two classroom components utilized by educators before grading practices occur: learning standards and student assessment (Brookhart, 2009; Cox, 2011). However both of these types of practices ultimately rely on grading, and the positive effects of these types of standard and assessment reforms may be limited because the majority of grading practices occurring within American public high schools are ineffective at their intended purpose (Brookhart, 2009; Guskey, 2015; Marzano 2000; Reeves, 2011).

The discrepancy between teacher grading practices and researcher recommendations has long been established. Studies documented the subjectivity

of teacher grades in the beginning of 20th century (Starch & Elliot, 1912, 1913a, 1913b), but at the beginning of the 21st century, teacher grading practices are not significantly different. Measurement specialists are very clear: student grades should represent only academic achievement, and the primary purpose of grades is to communicate that achievement to parents and other interested parties (Guskey, 2015; Marzano, 2000; Reeves, 2011). However, researchers continue to report that educators often disregard grading research (Brookhart, 1994; Reeves, 2011; Stiggins et al., 1989), as teachers continue to use varying combinations of achievement, effort, behavior, and attitude to best suit their practical needs in the classroom (Brookhart, 1991; Cross & Frary, 1999; Kelly, 2008; McMillan, 2003; Reeves, 2011). The subjectivity of grades is often evidenced by the low correlation between teacher-assigned grades and standardized test scores (Bowers, 2011; Brennan, Kim, Wenz-Gross, & Siperstein, 2001; Willingham, Pollack, & Lewis, 2002; Woodruff & Ziomek, 2004), which Brennan et al. found to range between 0.5 and 0.6 and amounts to just 24% to 36% similarity.

The teacher grading decision-making process may help to explain teachers' disregard for expert advice, and this process may bring us to the root of the grading problem. Research shows that teacher grading decision making is complex and often self-contradictory, as educators often utilize multiple factors when making grading decisions (McMillan, 2003; Randall & Engelhard, 2010; Shavelson, 1973; Zoekler, 2007). In the often-cited article "What is *The Basic Teaching Skill*," Richard Shavelson concisely explained the importance of decision making in the classroom by asserting, "*The basic teaching skill is decision making*" (p. 144). Researchers agree that for grading to be effective, teachers must ensure an effective decision-making process (Allen, 2005; Azeem, Afzal, & Majoka, 2010; Cauley & McMillan, 2000; Curren, 1995; McMillan,

2003; Parmigiani, 2012). This emphasis on grading decision-making has prompted some researchers to study this process with hopes of better comprehending the context of grading to provide teachers with future recommendations that are more pragmatic in nature (McMillan, 2003; McMillan & Nash, 2000; Smith, 2003).

Unfortunately, some school contexts are more challenging than others, and high-poverty schools encounter especially difficult circumstances. Student achievement data from around the United States display an academic achievement gap between students of low-income and high-income families (Borg et al., 2012; Dahl & Lochner, 2012; Duncan & Murnane, 2014; Heckman, 2006; Reardon, 2011; Sirin, 2005; Stull, 2013). Research shows that school-level impacts are significant in areas of high poverty, and even students who are not individually classified as living at or below the poverty line experience negative effects on achievement in schools of high poverty (Borg et al., 2012; Borman & Dowling; Silvernail, Sloan, Paul, Johnson, & Stump, 2014). In efforts to address the income achievement gap, research studies have investigated the effects of numerous factors on students of poverty such as cognitive development (Crook & Evans, 2014), school size (Coladarci, 2006; Fetler, 1989; Johnson, Howley, & Howley, 2002), participation in organized activities (Morris, 2015), parent involvement (Cooper & Crosnoe, 2007; Williams & Sánchez, 2011), instructional time (Rogers & Mirra, 2014), teacher quality (Borg et al., 2012), and out-of-school learning (Alexander, Entwisle, & Olson, 2001). Although many of these studies have helped to begin to explain the gap, the complexity of the problem leaves many additional factors to be discovered.

A number of studies have focused on the link between teacher grading practices and poverty. Teachers in high-poverty schools understand the challenges

with which they are faced, and they often attempt to use grading to address these issues. These studies have found that teachers in schools of high poverty utilize non-academic factors such as effort, behavior, and attitude in creating student grades more often than in schools of mid- and low-poverty. As a result, student grades in schools of high poverty are often inaccurate—usually inflated—and fail to accurately represent student academic achievement (Agnew, 1985; Cauley & McMillan, 2000; Howley et al., 2000; McMillan et al., 2002; Office of Educational Research and Improvement, 1994; J. Rothstein, 2004; Zwick & Himmelfarb, 2011).

Theoretical Perspective

This study employed both a critical pedagogy and teacher grading decision-making theoretical framework. The critical pedagogy framework is presented first, as it served as the foundational theoretical framework of the study and the broader lens through which the research problem was viewed in the context of society. The presentation of the critical pedagogy framework begins with an introduction to the overarching theory of social justice, and it proceeds to a look at the key components and authors of critical pedagogy. The presentation of the teacher grading decision-making framework follows, as it is a narrower lens through which the research problem was viewed in the context of the school classroom. This section begins with a general explanation of the framework and the two publications that led to the framework. Next is a more specific explanation of the framework's three domains and supporting literature. This section concludes with an explanation of the application to the study.

Critical Pedagogy Framework

In his book *Critical Pedagogy for Social Justice*, John Smyth (2011) demonstrated the importance of applying a critical lens to education to provide educators, students, parents, community members, and researchers with an understanding of the “winners and losers” in our education system and, just as importantly, steps that can be taken to approach social justice. This study attempts something similar, however, the scope is limited to teacher grading practices. Before explaining critical pedagogy and its specific role in this study, it may be helpful to provide an overview of the larger concept of social justice, as critical pedagogy is a means of working toward social justice in schools, and thus that work is the ultimate goal of this study.

Social justice. Social justice has become a popularly used term within the field of education by teachers and administrators alike, where it receives high use within mission and vision statements of schools and as a descriptor of teacher practices (Carlisle, Jackson, & George, 2006; Hytten & Bettez, 2011; Moule, 2005; North, 2006; Zollers, Albert, & Cochran-Smith, 2000). The recent popularity of the term is not without good reason: Many researchers call for educational leaders to actively work toward social justice in their school to improve the lives of students, students’ family members, and the entire school community (Jenlink & Jenlink, 2012; Shields, 2004; Smyth, 2012). However, the term is broad, multidimensional, and complex, and as a result it is often misunderstood (Gewirtz, 2006; Novak, 2000). Misunderstanding often reduces social justice to a rhetorical expression devoid of its rich meaning and great potential for positive effect (Gewirtz, 2006; Hytten & Bettez, 2011; North, 2006). Schools that have attempted to put the term into practice in an ambiguous form often experience resistance, confusion, tension, and conflict, among other

difficulties (Hyttén & Bettez, 2011; Moule, 2005; North, 2006). Any successful work toward social justice in education should begin with a firm understanding of the concept and a comprehensive framework to view and apply the concept (Hyttén & Bettez, 2011; Carlisle et al., 2006; Gewirtz, 2006).

Definitions of social justice range from broad to specific, but they all correspond in their complexity. Novak (2000) broadly defined social justice by explaining two essential characteristics of the term: It must be collaborative work toward justice, and it must be for the larger good, not the individual. Despite the emphasis on the collective good, Novak added that social justice is a virtue of individuals—one that is essential in a democracy. In contrast, Murrell (2006) provided a fairly specific definition of social justice in education as “a disposition toward recognizing and eradicating all forms of oppression and differential treatment extant in the practices and policies of institutions, as well as a fealty to participatory democracy as the means of this action” (p. 81). Some of the confusion over social justice, according to Gewirtz (1998), stems from its use as a synonym of distributive justice, which is primarily concerned with the equal distribution of goods within a society. This confusion seems to be perpetuated by the multidimensionality of social justice, as Gewirtz (2006) described three different dimensions—distributive, recognitional, and associational justice—and posited that although the three dimensions at times can be self-contradictory, they are appropriate in different contexts. North (2006) added some closure to the definition of the term by explaining that “even among competing theories of social justice, all scholars seem to agree on one thing: the liberal belief in citizens as free and equal persons merits celebration and protection” (p. 516).

Numerous models and frameworks have been created to apply social justice in schools. Hackman (2005) explained that successful application of social justice

education by educators in the field requires five essential components: content mastery, tools for critical analysis, tools for social change, tools for personal reflection, and an awareness of multicultural group dynamics. Carlisle et al. (2006) presented a framework to be used in schools consisting of principles of social justice: inclusion and equity, high expectations, reciprocal community relationships, system-wide approach, and direct social justice education and intervention. Additionally Hytten and Bettez (2011) significantly contributed to the topic by positing that when educators apply a social justice orientation, they may draw from one of many different discourses including “democratic education, critical pedagogy, multiculturalism, poststructuralism, feminism, queer theory, anti-oppressive education, cultural studies, postcolonialism, globalization, and critical race theory” (pp. 8-9). Like Gewirtz’s (2006) description of the multidimensionality of social justice, the various discourses often overlap in meaning and use and can also be contradictory (Hytten & Bettez, 2011; North, 2006). However, while Gewirtz (2006) explained that the use of a particular dimension of social justice often depends on the context, one’s social justice orientation rarely changes (Hytten & Bettez, 2011). Thus, it is essential for schools to concretely understand and agree upon their own interpretation of social justice before moving toward putting theory into practice. This study applied a social justice orientation based upon a critical pedagogy discourse.

Critical pedagogy. Critical pedagogy is a philosophy and method of analyzing the role of power within classrooms, schools, and society to empower the marginalized and thereby work toward social justice (Kincheloe, 2008; McLaren, 1997; Smyth, 2011). The philosophy specifically recognizes that education is *not* neutral and is an inherently political process. It emphasizes the importance of analyzing power structures at work within schools (Kincheloe,

2004, 2008), which include processes such as staff hiring, textbook adoption, novel selection, and curriculum creation (Apple, 2014; Nieto, 1996). Critical pedagogy posits that those in power influence and control the distribution of knowledge (McLaren, 1991). Kincheloe (2008) explained that “a critical dimension of critical pedagogy involves its understanding and use of knowledge,” and advocates of critical pedagogy, he argued, must “appreciate a variety of perspectives on the way knowledge is produced and deployed” (p. 10). Critical pedagogues assert that those in power display tendencies to oppress those who are not through various overt and covert means (Kincheloe, 2008), including using schools to reproduce values, skills, and world-views to maintain a stable society and a hold of power (Giroux, 1997). These decisions also include determining what knowledge is and is not disseminated to students (Apple, 2014).

Critical pedagogy engages in a power analysis from the perspective of those who are marginalized in schools often based on race, ethnicity, gender, class, language, sexual orientation, and neighborhood (Theoharis, 2007). This perspective allows both practitioners and students to gain a better understanding of the ways that schools often support and perpetuate the status quo and as a result further marginalize those outside of traditional majorities (McLaren, 1991; Smyth, 2011). Ultimately critical pedagogy exists for the purpose of alleviating human suffering in the form of individual and group oppression—something crucial for the sustenance of democracy (Kincheloe, 2004). Thus, critical pedagogy is used as a method of continuing and perpetuating democracy (Giroux, 2004; McLaren, 1991). Giroux (2011) began his book *On Critical Pedagogy* with an explanation that the philosophy is based upon the premise “that education is fundamental to democracy and that no democratic society can survive without formative culture shaped by pedagogical practices capable of creating citizens who are critical, self-

reflective, knowledgeable, and willing to make moral judgments and act in a socially responsible way” (p. 3). In *Affirming Diversity: The Sociopolitical Context of Multicultural Education*, Nieto (1996) added to this topic by explaining that multiculturalism *is* critical pedagogy, and a multicultural curriculum provides students with multiple perspectives of viewing the world essential to becoming active participants in a democratic society. Banks and Banks (1997) noted that the emphasis of diverse viewpoints by multiculturalism and critical pedagogy empowers students to develop critical judgment and decision-making skills important for democratic citizens.

Critical pedagogy largely originates from the work of Paulo Freire, and in particular, Freire’s (1996) seminal book *Pedagogy of the Oppressed*. In this work, Freire (1996) applied both political and educational theory to explain the oppression that occurs within society of the powerful over the powerless and steps needed to reach liberation. Freire emphasized the need for *praxis*—the combination of action and reflection—to accomplish freedom. Freire explained, “Liberation is a praxis: the action and reflection of men and women upon their world in order to transform it” (p. 79). Freire spoke out against traditional forms of education, which he called *banking education*, that authoritatively deposit knowledge into the minds of passive students and moves toward domestication. Rather, Freire stated that liberating education requires a partnership between teacher and student and extensive dialogue to reach *conscientização*, or a critical consciousness, to engage in praxis, and transform the world (Freire, 1996).

Application to the study. This study applied the critical pedagogy theoretical framework as a foundation of the study and a lens to view the study as a means to improving the lives of impoverished youth and working toward social justice. The lens provided a method to analyze the grading process from the

perspective of the student in a high-poverty school. Applying this perspective when analyzing schools of high poverty required questioning how the grading process affected these students. When analyzing schools of lower poverty levels, the perspective was used as a mode of comparison. Critical pedagogy was ultimately used as a lens to interpret findings on the impact of school poverty level on the teacher grading decision-making process.

Teacher Grading Decision-Making Framework

Building upon the foundation of critical pedagogy, this study applied a teacher grading decision-making framework (Figure 1) to the structure of the study to explain the complex processes that teachers utilize when grading. McMillan and Nash (2000) first created this framework to explain the processes involved in teacher assessment practices and teacher grading practices. Although this model was still in its early stages of development in the 2000 study, the article clearly articulated that both teacher assessment and grading practices were the result of a decision-making rationale that was influenced by a combination of factors internal to the teacher, factors external to the classroom, and classroom realities. In a follow-up article, McMillan (2003) further analyzed the original data and provided greater depth and description to the framework. However, the revised model failed to include grading practices as a result of the process, and instead it only included assessment practices (Figure 1). Despite this omission in the figure, McMillan's (2003) article included grading in the text description of the process, and it is this researcher's opinion that this 2003 framework is just as applicable to grading as it is to assessment. Thus, this teacher grading decision-making framework will be used as the second theoretical framework of the study.

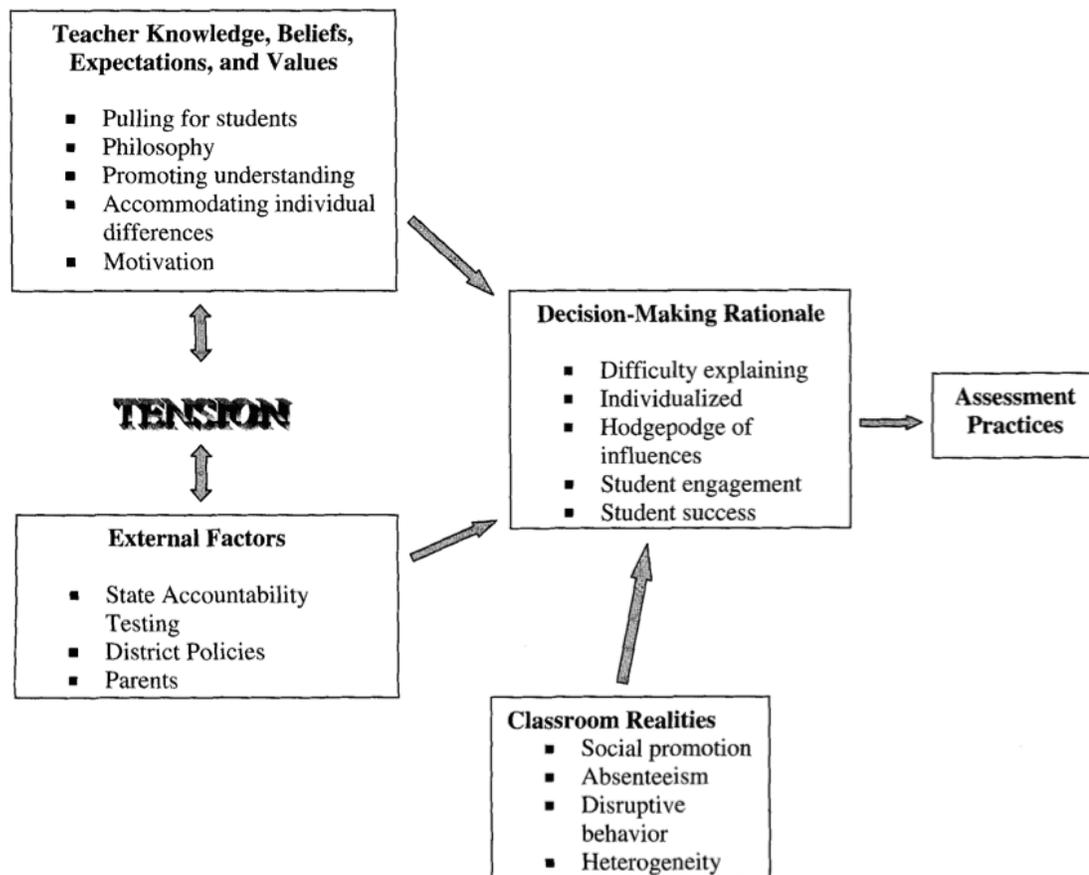


Figure 1. Teacher grading decision-making framework.

From “Understanding and Improving Teachers’ Classroom Assessment Decision Making: Implications for Theory and Practice” by J. H. McMillan, 2003, *Educational Measurement: Issues and Practice*, 22, p. 36. Copyright 2003 by J. H. McMillan. Reprinted with permission.

The framework consists of three domains: influences, rationale, and grading practices (Table 1). Three components comprise the decision-making influences: (a) internal factors (i.e., teacher knowledge, beliefs, expectations, and values; (b) external factors; and (c) classroom realities. These influences explain decision-making rationale, which directly leads to grading practices and finally student report card grades. A major factor within the framework is the tension created in balancing the influence of teacher internal factors with external factors, as this combination of influences is merged with the influence of classroom

realities to produce a decision-making rationale that ultimately leads to grading practices. It is this incongruence of the decision-making influences that explains grading practices that often differ from teacher education philosophy (McMillan, 2003; McMillan & Nash, 2000).

Table 1

Teacher Grading Decision-Making Framework by Domain

| Domain | Description | Components |
|--------|---------------------------|--|
| 1 | Influences of Grading | <ul style="list-style-type: none"> • Internal factors (i.e., teacher knowledge, beliefs, expectations, & values) • External factors • Classroom realities |
| 2 | Teacher Grading Rationale | -- |
| 3 | Teacher Grading Practices | -- |

Below is an explanation of McMillan's (2003) teacher grading decision-making theoretical framework. Each component is briefly described along with a number of contributing elements to provide a better understanding of the framework. A more detailed explanation of the studies that led to the creation of the framework along with supporting literature of the components is provided in Chapter 2.

Teacher knowledge, beliefs, expectations, and values. This component of the framework, described in this study as a component of Domain 1, comprises the major internal factors that teachers utilize in the decision-making process. McMillan (2003) described five primary elements that comprise this component: pulling for students, philosophy, promoting understanding, accommodating

individual differences, and motivation. To begin, teachers believe it is important to *pull for* student success, which means they often do anything they can to help their students to succeed, including utilizing grades to help students achieve success (Brookhart, 1993; Cizek et al., 1995/1996; McMillan, 2001, 2003; Randall & Engelhard, 2010). Teachers' philosophy of education is an additional significant contributor to this component, and it consists of teachers' most essential beliefs of teaching and learning. Teachers often use this philosophy to explain their grading decision making (McMillan, 2003). Grading literature has shown the importance of teacher values and beliefs in decision making and teacher actions (Allen, 2005; Clark & Peterson, 1984; Frisbie & Waltman, 1992; Senk, Beckmann, & Thompson, 1997; Thomas, 2014; Tomlinson, 2001b). McMillan (2003) also found that teachers use assessments to promote understanding, which is then reflected in grades in the form of achievement. To ensure fairness, which is very important to teachers when grading (Brookhart, 1993; Tomlinson, 2001a; Zoekler, 2007), teachers often accommodate individual differences in ways such as modifying assessments (Cizek et al., 1995/1996; McMillan, 2003). Finally, teachers believe that students must be motivated in order to highly achieve and actively engage in the classroom. Teachers often use grades to motivate students and facilitate engagement (Brookhart, 1997; Docan, 2006; Graham, 2005; McMillan, 2003; McMillan & Nash, 2000; Zoekler, 2007).

Classroom realities. Classroom realities, a second component of Domain 1, are the challenging elements of the classroom environment often out of the control of the teacher, but things they must address to ensure student engagement and ownership of learning (McMillan, 2003). McMillan and Nash (2000) described several pertinent components of the classroom reality factor: student home environment, student attitude and behavior, classroom heterogeneity, and

absenteeism. Teachers often consider students' challenging home environments when grading (McMillan, 2003; McMillan & Nash, 2000), while poor attitudes and inappropriate behaviors are often factors that impact grading—especially for remedial and lower level (below college preparatory) classes, as grading is often considered a tool for classroom management (Brookhart, 1993, 1994; McMillan, 2003; Wormeli, 2006). The heterogeneity of student ability level within a class may provide challenges to grading by pressuring teachers to inflate grades (Brookhart, 1993; Graham, 2005; Hargis, 1990; McMillan, 2003), and absenteeism plays a role in grading practices by requiring teachers to address missing grades, zeros, and make-up work (Farkas, Grobe, Sheehan, & Shuan, 1990; McMillan & Nash, 2000; Resh, 2009).

External factors. The third component of Domain 1, external factors, is the external influences on the classroom that do not originate from the student and are not under the control of the teacher. These factors include three major elements: state accountability testing, district grading policies, and parents. State standards and testing indirectly affect grading practices by influencing the content and assessments that teachers use in the classroom that are ultimately graded (McMillan, 2003; Shepard, 2008). District grading policies are shown to have minimal, but detectable influences of grading practices, and interpretations of district grading policies greatly vary (Austin & McCann, 1992; Frisbie & Waltman, 1992; McMillan, 2002; Randall & Engelhard, 2010). Finally, parents also have influence on grading, as teachers want to have justification to explain grades to parents to avoid conflict. This can affect the number of grades that teachers give and the objectivity of grades (McMillan, 2003; Waltman & Frisbie, 1994; Zoeckler, 2007).

Decision-making rationale. The second domain of the teacher grading decision-making framework is rationale, which refers to the logic and reasoning that teachers utilize when making grading decisions. The teacher grading decision-making framework explains that the three components of Domain 1—teacher knowledge, beliefs, expectations, and values; classroom realities; and external factors—are all contributing factors of teacher grading rationale (McMillan, 2003). However, McMillan and Nash (2000) found that teachers often have a difficult time explaining this rationale. Although the teacher grading decision-making framework displays a number of possible factors at work within this domain, the authors emphasized two consistent findings from their studies of teacher rationale for grading decisions: a wide range of criteria and professional experience (McMillan, 2003). Teachers believe that they should use multiple grading criteria to fairly grade students, and they often use a hodgepodge of factors (Brookhart, 1991; Cross & Frary, 1999; McMillan, 2003; Reeves, 2011). Additionally, professional experience is shown to be a significant factor in rationale, as teachers described their development of practice through experience (Frisbie & Waltman, 1992; McMillan, 2003; Reeves, 2011). This experience seemed to be their own personal experiences or informal learning from fellow teachers and completely different from training and professional development.

Application to the study. The critical pedagogy theoretical framework was applied to understand and analyze the differential application of the grading process in schools of different poverty levels. This framework aided the researcher in applying multiple viewpoints to understand grading phenomena, and most importantly to focus on marginalized students, who in this case were students in high-poverty schools. The teacher grading decision-making framework was a more practical framework that allowed the researcher to analyze grading as a

process, instead of a result. While a component of this analysis was the teacher internal processing, this was not the emphasis, and more sophisticated cognitive theories exist to explain this single component of the process. More specifically, this framework allowed the grading process to be broken down into three separate domains to be studied and analyzed by multiple poverty levels.

Significance of the Study

Grades are essential to our current educational system. Regardless of the intent of a teacher in assigning any course grade, Brookhart (1999) pointed out that the future meaning of *all* grades is assumed to be the achievement of a student in a particular course. Because grades form such an important part of students' current and future lives, it is essential that we ensure their integrity. As measurement and accountability in schools has been a major emphasis over the past 15 years, grading is more important today than ever before. During the recent standards-based reform movement, criterion-based standards and standardized and local assessments have been the major focal points (Brookhart, 2009), but grading reform has not yet been a point of emphasis for educators (Cox, 2011). It seems that it is only a matter of time until grading becomes a key issue for reformers. Perhaps even more importantly, this study aims to better the lives of a marginalized group: impoverished students. Improving grading practices may not immediately impact the lives of these students, but it does have the possibility of indirectly doing so by improving their education.

Overall, the grading literature base is extensive; however, it does exhibit deficiencies in several areas. Over the last 30 years, the literature is plentiful with recommendations for grading practices, documentation and analysis of the discrepancies between researcher recommendations and teacher practices, and more recently suggestions for reforms such as standards-based grading and

minimum grading. But overall the research is much less plentiful since the beginning of the 21st century, and there are only a small number of studies that attempt to explain the reasons for the decades of poor teacher grading practices. More specifically, grading literature is scarce in the area of grading decision making, and although poverty and the income achievement gap are well studied, there is a deficiency of research making the connection between grading and poverty. Upon thorough review of the literature in this area, there is a clear need for literature addressing the impact of school poverty level on teacher grading decision making.

The present study adds to the existing literature by connecting three components for the first time: grading, teacher decision making, and poverty. Addressing this gap in the literature can potentially benefit policymakers, researchers, administrators, teachers, and parents to better understand the current issues in grading and work toward solving these problems.

Purpose of the Study

The purpose of this study was to investigate the effects of school poverty level on teacher grading decision making. A transformative explanatory sequential mixed methods design (Creswell, 2014) was used in which critical pedagogy provided an overarching theoretical framework for the study. This lens was used because it provided a means for focusing on and analyzing the influences of school poverty level in the study. The study involved using both quantitative and qualitative data gathered sequentially. In the first, quantitative phase, a survey structured around a teacher grading decision-making framework was utilized to collect data from teachers in high schools throughout a single high school district to assess the relationship between school poverty level and teacher grading decision making. The second, qualitative phase was used to explain the

quantitative results and address the teacher grading rationale component of the decision-making framework. This exploratory section of the study applied a constant comparative analysis method to qualitative data from the survey and focus groups to explore the grading decision-making process of high school teachers at three different schools poverty levels.

Research Questions

The research questions of the study align to McMillan's (2003) teacher grading decision-making framework (Table 1, p. 14). The study utilized a single guiding research question: How does school poverty level affect the teacher grading decision-making process?

This guiding question was broken into three specific questions to address each of the three domains of the teacher grading decision-making framework:

1. How does school poverty level affect teacher grading practices?
2. How does school poverty level affect the influences of teacher grading?
3. What rationale do teachers of different school poverty levels use to make grading decisions?

Summary

Teacher grading practices are often ineffective, and high-poverty schools appear to use ineffective practices more often than low-poverty schools. The grading literature shows that teacher grading practices have been known to be ineffective in education for more than 100 years, yet they remain largely unchanged today. Inaccurate grades resulting from hodgepodge grading practices fail to accomplish their primary purpose of communicating student academic achievement. Studies have shown that high-poverty schools are more likely than

lower-poverty schools to produce inaccurate grades. This study utilized a critical pedagogy framework as a foundation to view and interpret results from the perspective of students in high-poverty schools, while a teacher grading decision-making framework was utilized as a way to better understand the grading process and analyze the different components of the process—both internal and external—by varying school poverty levels. This study is of particular significance because as far as the author is aware, it is the first to combine factors of grading, teacher decision making, and poverty in a study. Ultimately, the purpose of the study was to analyze the impact of school poverty level on the teacher grading decision-making process in order to inform researchers and educators of this particular problem and facilitate a method of solving this problem in the classroom.

Definition of Key Terms

Academic achievement – Teacher judgment of student success in a class over a period of time utilizing only student performance on academic tasks as evidence for the judgment.

Academic factors – Criteria utilized by teachers to create student report card grades derived from academic tasks (e.g., test scores, essay scores, or quiz scores).

Achievement gap – The difference in the academic achievement between a majority group and a minority group(s).

Assessment – A formative or summative method of determining a students' understanding of one or more academic topics.

Common practices – Teaching practices conducted commonly among teachers of the same class or subject.

Free or reduced price lunch (FRPL) – Meals provided to public schools through the National School Lunch Program. Students from families of income at or below 130% of the poverty line are provided with free meals, and students from

families of income between 185% and 130% of the poverty line are provided with reduced-price meals (www.fns.esda.gov).

Grades – Symbols utilized by teachers to represent student achievement on assignments, assessments, or report cards.

Grading – The process of a teacher scoring student assignments or assessments (Brookhart, 2009) *or* the process of a teacher assigning students a symbol to represent their achievement in a class on a report card for a marking period. Because this term is used in two possible ways, the context of its use will determine its meaning.

Higher-income students – Students from families of higher income levels, which in this study is defined as families above the poverty line traditionally thought of as middle and upper class. This term is synonymous with *higher-SES students*.

Higher-level academic classes – Classes of high academic rigor, which include honors, Gifted and Talented Education (GATE), and Advanced Placement (AP) classes.

High-poverty schools – Public schools in which more than 75.0% of students are eligible for FRPL.

Income achievement gap – The disparity in academic achievement between students from families of low-income and those from families of higher-income levels (Silvernail et al., 2014).

Lower-level academic classes – Classes of lower levels of academic rigor as compared to mid- and higher-level academic classes. In this study, the term is used synonymously with general level classes. These courses are designed to prepare students to directly enter the workforce, trade schools, professional training, or the military.

Low-income students – Students from families of low-income, which in this study is defined as families at or below the poverty line. This term is synonymous with *students of poverty* and *low-SES students*.

Low-poverty schools – Public schools in which 50.0% or less of students are eligible for FRPL.

Mid-level academic classes – Classes of average levels of academic rigor. In this study, the term is used synonymously with college preparatory (CP) level classes. These courses are designed to prepare students for college, but they are not as rigorous as honors, GATE, and AP classes.

Mid-poverty schools – Public schools in which 50.1% to 75.0% of students qualify for FRPL.

National School Lunch Program (NSLP) – The federal school meal program that provides students from families of income at or below 130% of the poverty line with free meals and students from families of income between 185% and 130% of the poverty line with reduced price meals (fns.esda.gov).

Non-academic factors – Criteria utilized by teachers to create student report card grades derived from non-academic tasks (e.g., ability, behavior, or effort).

South San Joaquin Valley – The area consisting of Kern, Fresno, and Tulare counties (Beacon Economics, 2014).

CHAPTER 2: REVIEW OF THE LITERATURE

This chapter provides a comprehensive review of literature on the impact of school poverty level on the teacher grading decision-making process. To do so, the first section of the review is devoted to the independent variable, poverty in schools. Following is a review of the literature on the dependent variables: grading in schools and the teacher grading decision-making process. The review culminates with literature on the link between poverty, teacher grading, and decision making.

The Impact of Poverty on Student Achievement

Students living in poverty face long odds in being successful in school. Tyrone Howard, author of *Why Race and Culture Matter in Schools: Closing the Achievement Gap in America's Classrooms* (2010), declared, "One of the most critical factors behind schooling disparities is socioeconomic status" (p. 2). Many researchers have found that these disparities arise much before children begin school (Evans & Rosenbaum, 2008; Heckman, 2006; Lee & Burkam, 2002; R. Rothstein, 2004). By the time school begins, not only do children of poverty need to learn the required standards for each class like other students, but these kids also must learn the prerequisite skills that they often lack, unlike children of families of higher income (Ready, 2010; R. Rothstein, 2004; Stull, 2013). A multitude of factors that act on students of families of low-income depresses the academic performance of these children, which results in an academic achievement gap between students of families of low-income and those of families of mid- and high-income (Alexander et al., 2001; Crook & Evans, 2014; Dahl & Lochner, 2012; Reardon, 2011; Sirin, 2005). This section of the literature review is organized into two sub-sections: literature devoted to describing and

understanding the income achievement gap and literature focused on explaining the factors that lead to the gap.

Understanding the Income Achievement Gap

The most common use of the term *achievement gap* refers to the disparity in academic achievement between African American, Latino, Native American, and some Asian (primarily Southeast Asian) students on the low-end and White and certain Asian American (primarily Chinese, Japanese, and Korean) students on the high-end (Howard, 2010). However, poverty, family income, and socioeconomic status (SES) all have a significant impact on education, and the achievement gap addressed in this study is the one between students from families of low-income and students from families of higher-income levels (Battle & Lewis, 2002; Heckman, 2006; Howard, 2010; Reardon, 2011; R. Rothstein, 2004; Sirin, 2005; Toutkoushian & Curtis, 2005; White, 1982).

The presence of an income achievement gap is supported by an abundance of evidence; however, the precise form of the independent variable that authors utilize in studies on the topic often varies. The common forms of the variable include *family income*, *school or family SES*, *school or student social class*, *student (or family) poverty*, and *school poverty level* (e.g., Borman & Dowling, 2010; Duncan & Murnane, 2014; Rogers & Mirra, 2014; Shuffleton, 2013; Sirin, 2005; Stull, 2013). Although these variables are similar and sometimes used synonymously, they have important differences. While family income is specific to money earned in a household, forms of the term *student (or family) poverty* often represent a compilation of data on household income, the number of people in the a household, and the poverty threshold (U.S. Department of Education). SES, on the other hand, was described by Hattie (2009) as one's place in society's hierarchy, and he explained that it is directly related to home resources. Hattie

described these resources, which are considered the main indicators of SES, as parent education, parent occupation, and family income. Sirin (2005) explained that school SES is usually measured as the proportion of students at a school that are eligible for free or reduced price lunch programs (FRPL) in a school year. Students are eligible for the National School Lunch Program if their family is at or below 130% of the poverty level, while students whose families are between 130% and 185% of poverty level are eligible for reduced lunch prices (Sirin, 2005; United States Department of Agriculture, 2013). This definition is also used for school poverty level (e.g., Cauley & McMillan, 2000; Rauschenberg, 2014; Rogers & Mirra, 2014).

Documenting the gap. In *Visible Learning*, John Hattie (2009) synthesized over 800 meta-analyses on student academic achievement, and four of these meta-analyses focused on the interaction between SES and achievement. This particular synthesis included 499 total studies and 176,915 participants, and results showed an overall effect size of $d = 0.57$. Cohen (1988) explained that effect sizes of $d = 0.2$ are considered small, $d = 0.5$ are considered medium, and $d = 0.8$ are considered large. However, Hattie revised Cohen's interpretations of effect sizes for educational outcomes, as he described that effect sizes of $d = 0.2$ are small, $d = 0.4$ are medium, and $d = 0.6$ are large. Thus, Hattie's results of $d = 0.57$ may be interpreted as large, or very near large, and they display the significant role that SES plays in student academic achievement.

One of the four meta-analyses used by Hattie (2009) to analyze the role of SES on achievement was the work of Sirin (2005), and it was of particular importance. The Sirin study included 58 articles published between 1990 and 2000 that studied the relationship between SES and student achievement. The combined samples included 101,157 students, 6,871 schools, and 128 districts.

Results showed a medium-to-strong relationship between SES and achievement with a correlation effect size of $r = 0.32$. As r values may be directly calculated from the d value, a correlation effect size of $r = 0.10$ is considered small, $r = 0.24$ is medium, and $r = 0.37$ is large, according to Cohen's (1988) interpretations. According to Hattie's revised education effect size interpretation, $r = 0.32$ may be interpreted as a large effect size. Additionally, Sirin found that this relationship increased from kindergarten ($r = 0.19$) through elementary school ($r = 0.27$) and into middle school ($r = 0.31$), but in high school, the relationship reduced to earlier levels ($r = 0.26$). Effect sizes also differed by school setting: In rural schools, he found an effect size of $r = 0.17$; in suburban schools, $r = 0.28$; and in urban schools, $r = 0.24$. Finally, the author found that the relationship between SES and student achievement was stronger for White students ($r = 0.27$) than for minority students ($r = 0.17$). Sirin explained that this was likely because minority students—regardless of SES level—are more likely to live in areas with greater educational risk factors as compared to White students.

In a chapter in the book *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances*, Reardon (2011) presented findings that as the discrepancy between high- and low-income families has increased over the past 50 years, so too has the income achievement gap. Reardon found that the income achievement gap was 30 to 40 times larger for children born in 2001 as compared children born in 1976. Additionally, the author reported that the income achievement gap was almost twice as large as the Black-White achievement gap. Contradicting Sirin's (2005) findings, this study found that the gap did not increase as students progress through school, and the increase in the gap was not due to parent education—this factor remained relatively constant during the time. Reardon explained that the increasing income achievement gap may be partially

attributed to an increase in the private investment in child learning and development by higher-income parents.

The Coleman Report. Researchers and educators have not always agreed that schools play a major role in the relationship between poverty and student achievement. In the “Equality of Educational Opportunity” study, today often known as the “Coleman Report,” Coleman et al. (1966) conducted a survey of educational opportunity in the United States as a component of the Civil Rights Act of 1964. The study was large in scope, with a data sample of 570,000 students, 40,000 teachers, and 4,000 principals. Although it was designed to address four different issues, the Coleman Report is most known for its attention to the comparison of Black and White public schools. The study, which was reported directly to former President Johnson and Congress, concluded that schools are extremely similar when controlled for student SES and also that differences between schools minimally explain differences in student achievement. The study concluded that schools could do little to impact the Black-White achievement gap and student background and SES were the greatest indicators of student achievement (Coleman et al., 1966).

Although well-known by the general public, the findings of the Coleman Report are frequently misunderstood (Alexander et al., 2001; R. Rothstein, 2004). Richard Rothstein (2004) explained that a few particular misunderstandings of the report prevent clear public discussion of the influence of social class on learning. To begin, the public often interprets the report’s results to be that schools have no effect on student learning. This interpretation is incorrect, as the authors explicitly state that schools do impact student learning, but that the school effect on achievement is small in comparison to the effect of student background and SES. Richard Rothstein (2004) described that the second significant misunderstanding

is the way in which the achievement gap is measured. Although it was once measured in terms of norm-referenced data, it is now based on criterion-referenced data, and the size of the gap is largely dependent upon the difficulty of the standards at each cut point. The third significant misunderstanding is the meaning of group average achievement. Richard Rothstein (2004) explained that there is a tendency to conflate group average achievement (e.g., average achievement in high-poverty schools) with the achievement of each individual within a group, which is completely erroneous.

Borman and Dowling (2010) conducted one of a number of studies that replicated the Coleman Report to find different results. In addition to replicating the study, the authors also applied an HLM model for further analysis, which was not available during the publication of the original report. In stark contrast to the Coleman Report, Borman and Dowling found evidence that schools have a significant impact on student learning, as 40% of the variability in verbal achievement scores was found between schools. Additionally, the authors found that school social class and racial/ethnic makeup were 1 $\frac{3}{4}$ times more influential on student achievement than individual student social class or race/ethnicity (Borman & Dowling, 2010).

Explaining the Income Achievement Gap

There are numerous factors that contribute to the income achievement gap. While each individual factor may explain little, collectively the known contributing factors explain a significant amount of the gap (R. Rothstein, 2008). Researchers primarily utilize two categories of factors to explain this gap: those related to school and those related to social class (R. Rothstein, 2004).

School factors. School factors are school-related elements that may directly or indirectly contribute to the income achievement gap. These factors

include, but are not limited to, individual school effects, learning time, teachers, school funding, and school size.

Individual school effects. Individual schools have been shown to play a significant role in the relationship between poverty and student achievement (Borman & Dowling, 2010). The importance of the issue is highlighted in Borman and Dowling's statement, "The question of how the poverty and minority concentration within a school affects a student's achievement outcomes above and beyond the effect of his or her individual poverty and minority status is at the core of the sociology of education" (p. 1239). In discussing the implications of the significant effect sizes he found between SES and student achievement, Hattie (2009) explained, "SES is more important at the school than at the individual level" (p. 63). Silvernail et al. (2014) conducted a study on school-level poverty and academic achievement in Maine schools and found not only a negative relationship between student poverty level and student achievement, but also a significant school-level effect. The authors asserted, "The level of poverty in a school is the single best predictor of average student performance" (p. 29). Results of the study also showed that in schools of higher levels of poverty, students who did not qualify for FRPL (the determinant of poverty in this study) scored lower than similar students in schools of lower poverty levels.

Toutkoushian and Curtis (2005) analyzed achievement data from all 73 public New Hampshire high schools and found three indicators of school SES—unemployment rate, parent income, and adult education—to account for more than 50% of the variation in student standardized test scores. The authors concluded that when schools are compared based on standardized test scores, states should consider utilizing multi-outcome measures that consider school SES levels to develop a value-added formula (Toutkoushian & Curtis, 2005). While individual

family poverty has been shown to have a negative impact on student achievement, schools with high concentrations of student poverty seem to magnify the negative effects of poverty (R. Rothstein, 2013). Additionally, a number of studies have found that the achievement of low-income students would increase if they moved to low-poverty schools (Borg et al., 2012).

Sirin (2005) warned that using aggregate SES data might introduce ecological fallacies, as studies occasionally extend individual-level inferences and conclusions to school-level inferences, and vice versa. To avoid these issues, the author explained, researchers must ensure to interpret results properly. Any findings between SES and student achievement at the school level do not imply the same interaction for each student within the school. Conversely, findings of a relationship between SES and student achievement for individual students in schools should not be used to explain differences between schools (Sirin, 2005). Zwick and Green (2007) cautioned that a reason that some studies display a low correlation between high school grades, SAT scores, and SES level is because of the confounding of within-school data and between-school data. The authors explained that when within-school data is pooled, stronger and more accurate correlations between these achievement measures and SES are displayed.

Learning time. School learning time is especially important for low-income students; however, these students often have significantly less learning time than students of higher SES levels. Rogers and Mirra (2014) studied in-school learning time and educational opportunity by surveying a sample of 783 California teachers from 193 high schools. The authors found that although all California schools allocate the same amount of time in their instructional calendars, available learning time across schools is inequitable. The authors posited that learning time in high-concentration poverty schools is much less than

in low-concentration poverty schools. Survey results supported this statement with findings that high-poverty schools experience excessive instances of teacher non-teaching duties (i.e., lunch supervision, janitorial tasks, clerical tasks, and covering for absent colleagues), teachers playing multiple roles, lost instructional days (i.e., due to tested days, special days, disrupted days, and absent teacher days), classroom interruptions, chronic conditions impacting classroom time (i.e., noisy classrooms, insufficient qualified substitutes, and dirty classrooms), and minutes of non-instructional time (i.e., due to delayed starts, routines, and interruptions). The authors concluded that less learning time translates into less educational opportunity for students of high-concentration poverty schools, thus adding to the income achievement gap (Rogers & Mirra, 2014).

Ready (2010) analyzed early school attendance and cognitive development across student SES levels using the math and literacy scores of 13,613 children from 903 public and private schools obtained from the Early Childhood Longitudinal Study, Kindergarten Cohort. Findings suggested that in kindergarten and first grade, low-SES students who regularly attended school displayed higher rates of literacy learning than those regularly attending high-SES students, thus narrowing the achievement gap. However, low-SES students were more likely to be chronically absent, and literacy development displayed a strong relationship with attendance for these students. Conversely, math development did not display a significant relationship with attendance by class (Ready, 2010). Ginsburg, Jordan, and Chang (2014) had similar findings in an analysis of the 2013 National Assessment for Educational Progress (NAEP) data, as they found a relationship between student attendance and NAEP performance, student SES and attendance, and student SES and NAEP performance. Students of low SES were found to be

more likely chronically absent and to have lower performance than students of higher SES.

Teachers. Borg et al. (2012) used the math and English achievement data of 15,552 fourth and fifth grade students in a single Florida urban/suburban school district to compare high-poverty and low-poverty schools in order to better understand the income achievement gap. In this study, the authors defined high-poverty schools as those in which more than 56% of students qualified for FRPL, while low-poverty schools were defined as those in which less than 56% of students qualified for FRPL. Results showed that high-quality teachers (as measured by years of experience and level of education), smaller class sizes, and low teacher turnover all helped to improve student achievement and were more common in low-poverty schools. However, the authors also found that 85-87% of test score variation was due to within-school student characteristics (Borg et al., 2012). Simon and Moore Johnson (2015) studied high teacher turnover rates in high-poverty schools by analyzing six studies on the topic through an alternative lens that focused on school context as the probable cause. Findings suggested that high turnover in high-poverty schools is likely due to poor teacher working conditions in these schools and not student demographics. The authors explained that in addressing the problem, administrators should work to improve the poor conditions instead of solely focusing on recruiting new teachers to fill the vacancies (Simon & Moore Johnson, 2015).

In the often-cited article entitled “The Pedagogy of Poverty Versus Good Teaching,” Haberman (1991) described common forms of pedagogy used in high-poverty schools and suggested education reform to address these structural deficits. The author listed 14 practices that constitute the essence of teaching in high-poverty schools:

- giving information
- asking questions
- giving directions
- making assignments
- monitoring seatwork
- reviewing assignments
- giving tests
- reviewing tests
- assigning homework
- reviewing homework
- settling disputes
- punishing noncompliance
- marking papers, and
- giving grades (Haberman, 1991, p. 291)

Haberman explained that these practices are not only to be blamed on teachers, but that students, parents, the community, and society play an equal role in the perpetuation of the practices by expecting them in high-poverty schools and viewing the practices as the essence of teaching. Further, the author asserted that these practices appeal to several different groups: those who did poorly in school, who rely on common sense instead of research and other evidence, who fear the poor and minorities, who have low expectations of the poor and minorities, and who are unaware of more successful forms of pedagogy (Haberman, 1991).

School funding. High-poverty schools are not without additional resources to aid in addressing the income achievement gap. Title I of the Elementary and Secondary Act provides additional federal funding for high-concentration poverty schools and local educational agencies (LEAs) (U.S. Department of Education). Puma et al. (1997) conducted a study for the Department of Education with a nationally representative sample of students in first, third, and seventh grades to determine the effects of Chapter 1 funding (now termed Title I) on student achievement. Results showed that Chapter 1 school funding did not close the achievement gap between advantaged and disadvantaged students, and although

they found that schools did make an impact on student achievement measured by standardized tests, individual student and family characteristics contributed the most to student variations in test scores (Puma et al., 1997). Carmichael (1997) studied the distribution of Title I funds to New York State public schools using a descriptive analysis. The author found that Title I funds were spread thin throughout the state to 98% of all districts and over 80% of all schools, including schools with 0 to 10% poverty levels. Additionally, because Title I funding eligibility requires each individual school site to have at least the same concentration of low-income families as the LEA as a whole, districts with high-poverty levels may have schools of even slightly lower poverty levels than the district average fail to qualify for Title I funding (Carmichael, 1997).

In a study that evaluated the impact of Title I funding on school finance and student academic achievement in New York City public schools, Van der Klaauw (2008) utilized a regression-discontinuity approach to analyze school-level data from 1993-2001. Results showed that Title I funding was unsuccessful in increasing student outcomes in high-poverty schools, and it even resulted in negative impacts in some years. Liu (2008) analyzed the allocation of Title I funds by the state, district, and school level and found that contrary to the idea that areas of greater poverty receive greater Title I funds, the opposite was often true. Liu found that at the state level, because Title I funding is dependent upon state per-pupil funding, states with the highest poverty rates received the least amount of funding per poor child. Conversely, district poverty level was positively associated with funding per poor child. However, because funding formulas are based upon the sheer number of students in poverty, Title I funds were greater in large districts as compared to mid-sized and small districts. Finally, at the school

level, Liu found that high-poverty schools did receive more Title I funding than low-poverty schools.

Hanushek (1997) summarized 377 studies over 30 years that studied the impact of school funding on student achievement. Results showed that there was no consistent relationship between school funding and student achievement. The author included a warning that simply adding funding to schools would not necessarily increase student achievement, and he also provided confirmation that the educational policy of the time was largely ineffective (Hanushek, 1997). Jaggia and Tuerck (2000) conducted a two-year study to assess the impact of school funding on Massachusetts school standardized test scores. Results of the study found school SES and past school performance to be the primary predictors of school achievement levels. Additionally, like Hanushek, the authors found that increasing funding for all schools did not necessarily increase student scores state wide, but they suggested that targeting funding and reducing class size in the low-performing (and mostly low-SES) schools would be a better investment of funds. Finally, the authors suggested using a value-added model to compare school achievement levels that consider school SES (Jaggia & Tuerck, 2000).

School size. Fetler (1989) conducted a study on the relationship between student dropout rates and poverty, school size, and student achievement utilizing 2 years of data from all California regular public high schools. Results showed higher dropout rates associated with higher school rates of poverty and larger school sizes. Johnson et al. (2002) also studied student achievement, school size, and poverty in another comprehensive study that used data from all schools in Arkansas. The researchers found a negative relationship between school size and student achievement, but the strength of the negative relationship depended upon the school SES level. Schools of high SES levels showed a weak negative

relationship, while schools of low SES levels displayed a strong negative relationship. Johnson et al. further found that smaller schools seemed to mitigate the negative relationship between poverty and student achievement, while larger schools and districts seemed to magnify these inequities. Coladarci (2006) used a sample of eighth graders in 216 Maine schools to test the hypothesis that the decreased relationship between student poverty and achievement in small schools is merely a statistical artifact as opposed to a significant school-level effect. Results were inconclusive, which ultimately led to a rejection of the above hypothesis and further support for the claim that small school size benefits students of poverty (Coladarci, 2006).

Social class factors. Social class factors include student class-related characteristics that can be directly or indirectly attributed to the income achievement gap. These factors include, but are not limited to, cultural capital, cultural interventions, brain development, summer learning loss, health factors, psychological factors, and parent and family factors.

Cultural capital. Sociologist and author Pierre Bourdieu is well known for his work on the concept of *cultural capital*, which he explained was a significant part of education and a contributor to societal inequity (Bourdieu & Passeron, 1990). In *Reproduction in Education, Society, and Culture*, Bourdieu and Passeron described cultural capital as the knowledge and unwritten rules most valued by the dominant class in society, and it was described as a large part of educational success. The authors explained that schools reward students who apply this type of capital; however, it is largely taught and passed down by families that possess it—primarily those of upper classes. Bourdieu and Passeron explained that as each generation passes on its cultural capital to each successive generation, society is reproduced and the gap between classes remains intact.

A number of studies have analyzed the effect of cultural capital in schools, such as Roscigno and Ainsworth-Darnell's (1999) analysis of National Education Longitudinal Survey (NELS) data, which found that cultural capital and educational resources moderately explained the income achievement gap. However, they also found that the positive academic effects of cultural trips and educational resources were lower for low-SES students as compared to higher-SES students (Roscigno & Ainsworth-Darnell, 1999). In addition to cultural capital, Bourdieu also explained that *habitus*, one's understanding of the world and their internalized understanding of the role and potential role that one plays in the world, is important for success in our society (Bourdieu & Passeron, 1990). Dumais (2002) stressed the importance of including habitus in studies of the impact of cultural capital on student achievement. Dumais' study, which he described as the first "to operationalize the concept of habitus alongside the concept of cultural capital" (p. 62), found that habitus had a bigger impact on student grades than cultural capital. The author found that low-SES students displayed lower participation rates in cultural activities than higher-SES students, but the study failed to assess the impact of habitus by class level, as it instead focused on gender differences (Dumais, 2002).

Instead of focusing on cultural capital, Farkas et al. (1990) chose to analyze the effects of a more general category of *cultural resources* on student academic success with the use of a cultural resources/social interaction model of gatekeeping. Using a sample of 486 seventh and eighth grade students from 22 middle schools in a large city school district, the authors found that after controlling for content mastery, student work habits and basic skills had a significant impact on student grades, as teachers rewarded those students who displayed desirable characteristics. These findings supported the idea that teachers

often act as gatekeepers in schools by rewarding desired behaviors and characteristics (Farkas et al., 1990).

Yosso (2005) utilized a critical race theory (CRT) perspective to challenge the traditional view that cultural capital is the social knowledge and skills attributed to those of the dominant class—a view that implies that the lower class and communities of color have a deficit of cultural and social capital and even a deficit of culture. While applying CRT, the author rejected the traditional White middle-class culture as a frame of reference and abandoned the cultural deficit view. Instead, the analysis of communities of color, which largely included those of lower class, began with an assumption that non-White cultures also possess many strengths that may be viewed as assets. Yosso concluded that communities of color employ numerous forms of cultural capital such as aspirational, navigational, social, linguistic, and familial capital (Yosso, 2005).

Cultural interventions. A number of programs have sought to address the income achievement gap by means of cultural intervention; however, this is a topic of significant criticism (Bomer, Dworin, May, & Semingson, 2008; Shuffleton, 2013; Smyth, 2012; Evans-Winters & Cowie, 2009). Among the most popular of these programs are Ruby Payne's aha! Process, the Harlem Children's Zone (HCZ) charter schools and community services, and the Knowledge Is Power Program (KIPP) charter schools (Shuffleton, 2013). Ruby Payne's work is primarily presented in her book *A Framework for Understanding Poverty*, which was designed as a resource to provide educators with a method of systematically understanding the characteristics of students and families in poverty and as a tool to provide specific strategies to enable educators to more effectively meet the group's needs (Payne, 1996). The HCZ is a non-profit organization serving primarily minority and low-income students from birth through college graduation

in 97 blocks of Harlem, New York. The comprehensive system combines over 20 community programs with charter schools employing a strong academic focus that utilize high standards and expectations, high quality teachers, and extended school hours to produce high-achieving students (Dobbie & Fryer, 2011; Yeh, 2013). KIPP Academies are high-achieving charter schools located throughout the country that serve low-income students in urban areas. These schools emphasize character education and high academic standards while utilizing high parent involvement, extended school days and hours, and highly motivated teachers to ensure student success in middle school, high school, and college (R. Rothstein, 2004; Yeh, 2013).

In an article that criticized the three popular cultural intervention programs, Shuffleton (2013) promoted civic friendship as an ethically legitimate alternative educational intervention. The author applied a social justice lens when analyzing Ruby Payne's *aha! Process*, KIPP, and the HCZ, and she explained that applying cultural interventions wrongly begins with a premise that the culture of those in poverty is somehow deficient. More specifically, Shuffleton asserted that teaching and instilling middle- and upper-class cultural values within impoverished students is a case of authority figures using their power to promulgate their worldviews to those most vulnerable. In *Pedagogy of the Oppressed*, Paulo Freire (1996) described cultural interventions as a form of oppression performed by middle and upper classes on those of the lower class. He claimed that this oppression would end only when those in power stopped seeing the oppressed as abstractions and instead sought to understand them on an individual level (Freire, 1996).

Ruby Payne's work continues to receive both high praise and condemnation (Bomer et al., 2008; Gorski, 2008; Shuffleton, 2013; Smyth, 2012; Evans-Winters

& Cowie, 2009). Shuffleton described Payne's *aha! Process* as extremely flawed, as it is based upon definitions of poverty and culture that are at odds with popular research. Additionally, the author claimed that Payne merely reinforces popular stereotypes of the poor that results in lowered teacher expectations of poor students (Shuffleton, 2013). Bomer et al. devoted an entire paper to analyzing Payne's book, and they were equally critical of the work. The authors emphasized Payne's application of *deficit thinking*, defined as the belief that "students who struggle or fail in school do so because of their own internal deficits or deficiencies" (Bomer et al., 2008, p. 2523), to explain the income achievement gap. Bomer et al. explained that deficit thinking is not only morally wrong, but also unsubstantiated in the literature and likely detrimental to future success of children in poverty.

Additional critiques of Ruby Payne's *aha! Process* come from Smyth (2012), who used the terms "damaging, thin, demeaning, [and] pathologizing" (p. 11), among other adjectives, to describe the work. Gorski (2008) reviewed the published critiques of Payne's work and synthesized them into "eight elements of oppression" (p. 132) used by Payne: an uncritical and self-serving "scholarship," the use of the concept *culture of poverty*, a prominence of stereotypes, an application of deficit theory, a void of discussion on classism, the denial that poverty is about race, the application of paternalism, and a use of compassionate conservatism. Finally, although Evans-Winters and Cowie (2009) credited Payne with generating much-needed dialogue on the topic and filling the void of addressing students in poverty, they were critical of Payne's use of age-old stereotypes of the poor and her emphasis on changing culture to increase academic success—a strategy that they claim is ineffective and an oversimplification.

Researcher analyses of HCZ and KIPP are not as severely critical as those of Payne's program, as analyses of the former programs provide a greater balance of praise and critique (Dobbie & Fryer, 2011; R. Rothstein, 2004; Shuffleton, 2013; Yeh, 2013). Dobbie and Fryer found HCZ charter schools to be so effective in closing the Black-White achievement gap that they asserted that the schools were sufficient for the task without the use of expensive and often-controversial community programs. Richard Rothstein (2004) acknowledged that KIPP academies are successful in producing highly achieving students, but he claimed they accomplish this by utilizing factors that average schools across the country simply cannot, such as highly motivated parents, energetic teachers with duties that include home visits, young teachers who earn much less than veteran teachers, and extended school days and academic calendars.

Shuffleton (2013) criticized HCZ and KIPP by explaining that when the programs teach and emphasize middle- and upper-class cultural practices to lower-class students as a model of educational success, they are misinterpreting Bourdieu's social theory and are instead implying that lower-class cultural practices are inferior to middle- and upper-class culture. Yeh (2013) analyzed the two programs to assess their potential for comprehensive implementation throughout the United States. Conclusions showed that the programs' current reliance on the highest levels of teachers coupled with high teacher attrition rates would render the programs virtually ineffective at increasing student achievement if all schools implemented them (Yeh, 2013).

Brain development. A number of studies have shown that poverty has a significant negative impact on child brain development. Crook and Evans (2014) found that the income achievement gap could be at least partly attributed to children's executive functioning—specifically their ability to plan. The authors

concluded that early childhood poverty is a predictor of fifth grade math and reading achievement and this relationship is mediated in part by student planning skills. Evans and Rosenbaum (2008) helped to explain the relationship between family income and brain development with their finding that child self-regulation of emotion and behavior serves as a mediator between family income level and child cognitive development. An additional negative effect on cognition and I.Q. levels includes exposure to lead paint, which Richard Rothstein (2004) explained is more likely to occur in children living in poverty because of their increased likelihood of living in older buildings (R. Rothstein, 2004).

Children who grow up in poverty are also often exposed to high levels of stress that can have damaging effects on brain development and child learning (Center on the Developing Child at Harvard University, 2010; Committee on Psychosocial Aspects of Child and Family Health, Committee on Early Childhood, Adoption, and Dependent Care, & Section on Developmental and Behavioral Pediatrics, 2012; Evans, Brooks-Gunn, & Klebanov, 2011; National Scientific Council on the Developing Child, 2005/2014; Shonkoff, 2010). A report from the Center on the Developing Child at Harvard University (2010) documented the potential negative impact of adverse early-childhood environments on brain development, which are often experienced by those living in poverty. The authors explained that the stresses of poverty can lead to changes in brain architecture and neurobiological systems, which may result in changes in the body's stress responses and ultimately increased risks of disease and shortened lifespans (Center on the Developing Child at Harvard University, 2010). A similar paper from the National Scientific Council on the Developing Child (2005/2014) explained that the combination of a lack of protective relationships and early childhood adversity may result in childhood toxic stress. This type of stress, defined as a "strong,

frequent, or prolonged activation of the body's stress management system" (p. 2), can result in a greater development in the stress-related portions of the brain, while the reasoning, planning, and behavioral portions may remain underdeveloped (National Scientific Council on the Developing Child, 2005/2014).

Summer learning loss. Despite the fact that schools often reduce the income achievement gap during the school year, the gap widens during the summer, as children of mid and high SES are often in conditions to facilitate out-of-school learning, while children of low SES are often in the opposite conditions (R. Rothstein, 2004). In a study analyzing academic achievement data of Baltimore public school students of low, mid, and high SES, Alexander et al. (2001) attributed a major factor in the income achievement gap to differences in summer learning. The authors found that high-SES students consistently displayed significant summer achievement gains in math and reading (although lower than in the school year), while mid-SES students often had small gains, and low-SES students averaged no gains. While most low-SES students displayed stagnant reading and math summer learning, a subset of these students regressed in achievement over the summer. These results imply that students of higher SES have homes and communities that facilitate learning time away from the school, while low-SES students often do not. Additionally, the fact that low-SES students generally display achievement gains during the school year but not during summer break implies that schools make important impacts on low-SES student achievement despite mitigating factors such as low family income and low parent education level (Alexander et al., 2001). Alexander et al. explained, "Schools do matter, and they matter the most when support for academic learning outside school is weak" (p. 183).

Ready (2010) found similar results for summer learning loss using data from 13,613 early elementary students from 903 public and private schools retrieved from the Early Childhood Longitudinal Study (ECLS-K). The author studied the summer between kindergarten and first grade and found that mid-SES students made no gains, while high-SES students made gains in literacy and low-SES students regressed in achievement. Gershenson (2013) studied differences in child and parent summer time use by SES level utilizing two different time-diary surveys. Results showed that children in low-SES families watched almost two more hours of television and engaged in 12 fewer minutes in conversation with adults as compared to children in higher-SES families, while differences in parent time use were marginal. The author posited that this difference in child summer time use was a potential contributor to the income achievement gap.

Health factors. Children in poverty are more likely to suffer from poor health, and this has a strong negative impact on academic achievement (R. Rothstein, 2004). As compared to children of higher SES, low-SES children suffer from poorer vision, hearing, and oral health; a greater frequency of asthma; less frequent medical care; poorer nutrition; higher exposure to second-hand smoke; and a higher frequency of alcohol in the womb (R. Rothstein, 2004, 2013). A 2010 National Scientific Council report warned that negative experiences in children's lives—similar to those experienced by many in poverty—potentially produce *biological memories*, defined as life experiences built into our bodies that create detrimental effects on children's physical and mental health throughout their lives (Center on the Developing Child at Harvard University, 2010). The report also explained that low birth weight, which is more common in children in poverty, leads to significant increased risks of heart disease, diabetes, and obesity (Center on the Developing Child at Harvard University, 2010). As mentioned

above, chronic stress, which is more common in children of poverty, has a strong relationship with poor health (Evans et al., 2011).

In a study that used a nationally representative sample of 17,219 kindergarteners from the Early Childhood Educational Survey (ECLS), Hair, Halle, Terry-Humen, Lavelle, and Calkins (2006) analyzed school readiness by student socio-demographic factors. The researchers found that cognitive, language, and social/emotional skills along with good health were important factors for incoming kindergarten students for high student achievement in math and reading by the end of first grade. Additionally, children with multiple socioeconomic disadvantages were found more likely to display a risk profile, which indicated lowered academic achievement, while students who displayed strengths in all of the factors were mostly from advantaged backgrounds. Among the implications of the study, Hair et al. explained that child health and social/emotional skills should be increasingly emphasized in the assessment of child school readiness, and community investments in these areas would likely lead to increased child school readiness.

Psychological factors. Maslow's (1943) seminal paper "A Theory of Human Motivation," a paper often referenced for its presentation of Maslow's *hierarchy of needs*, described five essential needs that motivate human behavior: physiological, safety, love, esteem, and self-actualization. Maslow posited that these needs are sequential and hierarchical, as a person is incognizant of each subsequent higher-level need until the previous need is fulfilled. The first need, the physiological, refers to basic physiological factors such as food, water, health, shelter, and clothing. The second need, safety, refers to the state of being and feeling physically safe and stable. The third need is love, and this refers to the need for strong relationships that produce love, affection, and belonging. The

fourth need, esteem, refers to one's need for confidence, esteem, and respect both from one's self and from the world. Maslow explained that this need comes from a desire for strength and also for reputation and prestige. The final need is self-actualization, which Maslow (1943) explained is the need for fulfillment and actualized potential.

In one of the many articles that apply Maslow's *hierarchy of needs* to contemporary issues, Prince and Howard (2002) applied the framework to the study of children in poverty. The authors discussed that impoverished children often have major deficiencies in the basic physiological, safety, and love needs. Prince and Howard explained that children in poverty often live in households that are void of adequate healthy food, lack medical attention, live in neighborhoods that are dangerous, and are often without the affection from parents and family more common in higher-level income families. The authors also explained that children of poverty often lack self-esteem, which can be developed through the experience of successes and self-actualization and often relies on past success in education. The central idea of the article was that many children of poverty lack the basic needs that are taken for granted by those of higher SES levels, and these needs must be addressed to ensure a just society. On a similar topic, Henwood, Derejko, Couture, and Padgett (2014) applied Maslow's *hierarchy of needs* as a theoretical lens to investigate the experiences of 63 formerly homeless adults with serious mental illness. Results showed that participants were more likely to address self-actualization if their basic needs were *not* met, while a more complex relationship between basic needs and self-actualization was also found (Henwood et al., 2014).

Parent and family factors. In a study of 489 inner-city families, Cooper and Crosnoe (2007) failed to find a significant relationship between parent

involvement and economic disadvantage. Equally interesting, they found a strong positive relationship between parent involvement and child academic orientation for economically disadvantaged families, while families of middle and upper class showed a negative relationship between parent involvement and child academic orientation (Cooper & Crosnow, 2007). In a similar area of study, Chin and Newman (2002) studied working poor families in New York City and found that the large amount of time required of parents to earn an income directly resulted in limited time available to be involved in their children's education.

Stull (2013) studied the effects of family SES on student achievement using a nationally representative sample of approximately 22,000 students from the ECLS. Results showed that family SES had a strong positive relationship with student achievement, and the percentage of parents expecting their children to earn at least a Bachelor's degree increased with family SES. Additionally, the percentage of high-SES parents expecting their low-achieving children to earn at least a Bachelor's degree was higher than the percentage of low- and middle-SES parents expecting their high-achieving children to earn at least a Bachelor's degree (Stull, 2013).

Richard Rothstein (2004) reported that high student mobility—large number of instances in which students change schools—decreases achievement for both the individual student and the schools that experience high levels of student mobility. The author explained that high rates of mobility among lower-class children are caused by a lack of affordable housing in urban areas, family breakup, and parent unemployment (R. Rothstein, 2004). South, Haynie, and Bose (2007) studied the reasons for increased dropout rates among students that display greater school and residential mobility. The study utilized a sample of 8,519 students 14 or older from the National Longitudinal Study of Adolescent Health, and results

showed that mobile students were approximately twice as likely as non-mobile students to drop out of school. These findings were explained by a number of reasons: compared to non-mobile students, mobile students had smaller networks of friends; the friends of mobile students displayed lower levels of academic achievement; and the mobile students were less of a central figure in their own social networks. Additional causes of increased dropout rates were that the mobile students displayed lower levels of achievement and were less involved in their schools (South et al., 2007). Hanushek, Kain, and Rivkin (2004) also found that school turnover leads to reduced gains in achievement for all involved parties—including mobile students and all students in the new school. Additionally, the authors found that these negative effects were strongest for low-income and minority students.

Morris (2015) studied the impact of child participation in organized activities (OAs) on academic achievement using data from the Educational Longitudinal Study of 2002. The author analyzed math achievement scores of 12,893 students of different social classes in both tenth and twelfth grades. Although Morris acknowledged that lower-class parenting styles are less likely to encourage children to be involved in OAs, results showed that disadvantaged students received significant academic improvement from the participation in OAs, while more advantaged students did not (Morris, 2015).

Finally, Zhan (2006) studied the link between parent assets and achievement, expectations, and school involvement. The author analyzed data from the National Longitudinal Survey of Youth, and after controlling for parent income and other parent factors, results showed a positive relationship between parent assets and child achievement, parent expectations, and school involvement. Parent expectations were found to partially mediate the relationship between

parent assets and student achievement. Zhan explained that including data on parent assets could increase the general understanding of the impact of family economic factors on student achievement.

Grading in American Public Schools

The literature on grading practices in the United States is extensive, and it dates back more than a century. This section of the literature review is divided into two sub-sections: *a history of grading practices and concerns* and *grading today*. The section begins with *a history of grading practices and concerns* by exploring grading studies and articles from the distant past to gain an understanding and appreciation for the grading practices with which teachers currently employ. After establishing this historical perspective, the second half of the section, *grading today*, is devoted to the literature on researcher recommendations for best practices, a description of the grading practices utilized by teachers today, and an exploration of current grading issue.

A History of Grading Practices and Concerns

Before one considers contemporary issues in grading, attention should first be given to lessons from the past. To study the history of grading practices in this country, we must look first to the university level, as it was here that education was first organized, educators first evaluated student learning, and we find the first records of grading practices in the United States. As the review progresses through the roughly 350 years of grading history in this country, patterns emerge in the form short periods of rapid change and also long stretches of continuity. Interestingly, many issues in grading that we currently face were also issues addressed in the past.

America's first grades. Although grading today often lacks objectivity, practices are increasingly subjective the farther one looks back in history (Guskey, 1994; Kirschenbaum et al., 1971; Smallwood, 1935). The grading traditions that continue today have their roots in the 17th century, as Oxford and Cambridge scholars brought the very first practices over the Atlantic to the American colonies between 1630 and 1641. Although few records exist from this time, educators at Harvard University (then known as Harvard College), which was founded in 1636, were the first evaluators of American students, and the first evidence is in the form of a public oral examination occurring before the awarding of a degree in 1646 (Brookhart, 2009; Smallwood, 1935). These grading practices continued at Harvard until 1775 and still longer at Yale University (formerly known as Collegiate School, then Yale College) and the College of William and Mary (Smallwood, 1935). Some of the first modern grading practices in our country began at Yale in the late 18th century with the advent of measurement scales. In 1785 the president of Yale, Ezra Stiles, described four grading categories used in an examination as *Optimi*, *Second Optimi*, *Inferiores* and *Peiores* (Brookhart 2009; Durm, 1993; Smallwood, 1935). This description displays the first use of grades to sort students, but it is apparent that grading by use of numbers or letters had not yet evolved. By 1800 educational leaders realized that evaluation was a necessity, and between 1800 and 1840, Harvard and Yale moved to written exams (Smallwood, 1935). It was during this period that Yale was also one of the first American colleges to develop a grading scale composed of numerical values, as it began using a 4.0 scale in 1813 (Brookhart, 2009; Durm, 1993; Smallwood, 1935). This is likely the foundation of the same 4-point grading system still used today by high schools, colleges, and universities (Durm, 1993).

Harvard did not begin using a numerical scale until 1830, and unlike Yale's 4.0 scale, Harvard utilized a scale of 20. Later in 1837, Harvard began using a 100-point scale in mathematics and philosophy (Smallwood, 1935). Sometime after its use of the 4.0 scale, Yale changed to a 9.0 scale, but upon request of the faculty in 1832, the school reverted to the 4.0 scale (Smallwood, 1935). Other schools were using numerical scores for grading around this time, including the University of Michigan, which changed from a numerical system to a pass-no-pass system in 1851 (Durm, 1993). Increasingly modern grading scales were used in 1877 at Harvard with the grading of student ability on a 100-point percentage scale and then ranking students into one of six possible divisions. Soon after was the first evidence of letter grades, as records of a student at Harvard receiving a "B" is found from 1883 (Smallwood, 1935). In this same year Yale first presented the idea of examination by individual professors instead of by the university. It was also around this time that the University of Michigan first attempted to banish exams all together, but later settled on providing each professor with the choice of whether or not to use them (Smallwood, 1935). In 1898 Mount Holyoke College (formerly known as Mount Holyoke Female Seminary) combined multiple grading systems to produce a six-category system similar to one often seen today: A (95-100%), B (90-94%), C (85-89%), D (80-84%), E (75-79%), F (Failed) (Smallwood, 1935).

Early grading in public K-12 schools. The early grading practices used in colleges and universities from the 1600s until the early-to-mid 1800s became the foundation for grading practices in the common school in America (Brookhart, 2009). Prior to this time, there was little use for grading in what we now know as K-12 schools. Rich students had personal tutors, and poor students did not attend school at all. Students who fell into the early working class may have attended

subscription schools, but examinations given in these schools often produced descriptive results that were specific to the learning needs of each individual student (Brookhart, 2009). For all other children, education rarely exceeded elementary school, which occurred in a one-room schoolhouse (Kirschenbaum et al., 1971). Any judgment of student learning was done descriptively by skill. In these times prior to the common school, it was the teacher who was responsible for ensuring the learning of all students, and to a large extent the effectiveness of the teacher was measured based upon the learning of the students. However, the primary responsibility for learning shifted from teacher to student with the advent of common schools (Hargis, 1990).

It was around 1840 that the number of common schools underwent significant growth in the United States, and the time also marked the first reported use of report cards (Brookhart, 2009). Between 1850 and 1870, the number of children educated in the country went from 13% to 20%, and this rapid increase led to increased organization in schools by age and grade level. Between 1870 and 1910, the number of public high schools went from 500 to 10,000, and student enrollment from elementary school to high school went from 6,871,000 to 17,813,000 (Kirschenbaum et al., 1971). By 1890 as many as 69% of the American population between 5 and 17 was in a common school, and percentage grades became the most popular form of grading (Brookhart, 2009). With these drastic increases in student populations, schools quickly evolved formal grading practices. High schools began commonly using 100-point percentage scales as a way of efficiently assessing student learning by the end of the century (Hargis, 1990). These first grades were primarily used to compare students for employers and as a way to screen applicants for colleges (Kirschenbaum et al., 1971).

As the new century began, major changes were in store for grading practices in schools across the country. Although grading on a 100-point percentage scale was initially accepted without major issue, by the first decade of the 20th century, this grading method was being widely questioned and criticized (Cureton, 1971). In 1912 and 1913 Starch and Elliott published three seminal studies that questioned the validity of teacher grading practices. Starch and Elliot (1912, 1913a, 1913b) demonstrated the great variability in grading practices by investigating the grading of identical English, math, and history papers by hundreds of teachers. In their findings the authors stated, “It is almost shocking to a mind of more than ordinary exactness to find that the range of marks given by different teachers to the same paper may be as large as 35 or 40 points” (Starch & Elliot, 1912, p. 454). These papers seemed to ignite the discussion on the subjectivity of grading across the country. In a literature review of grading practices in the early 20th century, Ayer (1933) explained that many researchers at the time began to make significant modifications to percentage marking systems. These changes included discussions of grading standards; detailing of effort, attitude, preparation, and achievement in each grade; and moving to ranking systems or normalized curves. In an astonishingly insightful endnote to the review, Ayer noted, “The chief values of these marking systems rest upon the objective clarity with which they are understood by teachers, pupils, and parents and the degree to which they rest upon valid and reliable instruments” (pp. 203-204).

Additional concerns about the reliability of percentage grades moved schools into a use of standardized testing and normalized grading curves (Cureton, 1971). The creation of the normalized grading curve, in practice often known as “grading on a curve,” is credited to Meyer (1908) in a paper entitled “The Grading

of Students.” Meyer attempted to address popular concerns about grade distributions with the advent of the normal distribution curve by ensuring a fixed proportion of students in five different categories of the grading distribution: 50% in the middle, 3% excellent and 3% failures, and 22% inferior and 22% superior (Cureton, 1971). Although the use of the normal curve, or bell curve, was later corroborated by evidence showing that intelligence scores fell approximately into a normalized distribution curve (Middleton, as cited in Guskey, 1994), concerns about the practice soon grew in the field of education in the 1930s. Despite these concerns, this method of grading continued to see popular use in high schools for the next 40 years, as normal-curve grading was often seen as useful for college selection purposes (Brookhart, 2009).

By 1920, the use of letter grades first became popularized as a method to solve problems of reliability in percentage grades (Brookhart, 2009). Throughout the 1920s, educators increasingly pushed objectivity in grades, and along with World War I came the rise of standardized testing. Increasingly more schools adopted a 5-point grading scale as an alternative to the 100-point scale. Additional experimentation with grading continued during the time, including the first known instance of the mastery approach, in which students could not move on to a new skill until they demonstrated mastery of the original skill. During the 1930s grading reform moved in two polar directions: one toward no grades and another toward increasingly objective grades. Those in favor of no grades advocated for individualized instruction, whole-child growth, and increased parent communication, while those for objective grades continued to push for increased standardized testing that aligned with grades (Kirschenbaum et al., 1971). Similar debates over grades continued in the field over the next 30 years without reaching any consensus.

A shifting of the pendulum. By the end of the 1940s, more than 80% of schools had some type of 5-point grading system, largely because of administrative ease and university satisfaction. This time period also brought the rise of concern over utilizing grades for instructional purposes. Debate over objectivity, grading curves, grading criteria, standards, and validity continued into the 1960s without any substantial changes (Kirschenbaum et al., 1971). The pressure of protests throughout the 1960s, which included pressure by students to eliminate grades, persuaded universities to experiment with descriptive 3- and 4-point scales with categories such as “honors,” “high pass,” “pass,” and “fail”. However, these changes did not last long, and schools soon shifted back to the systems that we continue to see today: 4-point numerical scales and 5-category letter-grade systems (Kirschenbaum et al., 1971). These topics bring us to the 1980s, which represent the beginning of modern era grading practices.

Grading Today

Over the past 30 years, a great deal of research has been conducted on grading practices in high schools across the country. The crux of the concern has been the discrepancy between researcher grading recommendations and teacher grading practices. To better understand and recommend future reform measures, it is helpful to fully understand current researcher recommendations along with the various current teacher grading practices.

Five common purposes of grading. In his most recent book *On Your Mark: Challenging the Conventions of Grading and Reporting*, Thomas Guskey (2015) suggested the first step to addressing problems in grading is to clarify the purposes of grading. Unfortunately, this is often a difficult and challenging task. With this issue in mind, five major purposes of grading are described below. These purposes consist of a conglomeration of the major purposes of grading

reported in the literature for all interested parties, including parents, students, teachers, administrators, employers, and colleges.

Student achievement. The single greatest purpose for student grades is to communicate student academic achievement to students, parents, and other interested parties (Guskey, 2015; Marzano, 2000; Reeves, 2011; Wormeli, 2006). This purpose is the least controversial of all stated purposes of grading and is the only purpose to garner support from all interested parties. It is this single purpose that researchers recommend teachers to utilize when grading (Guskey, 2015; Marzano, 2000; Reeves, 2011; Wormeli, 2006).

Administrative purposes. Grades are widely used by schools and universities for administrative purposes. Although this purpose may not immediately come to mind for teachers, parents, and students, grades have served as a means to rank, matriculate, and retain students since their first use in this country (Guskey, 2015; Marzano, 2000). This purpose also serves as a factor in admission in post-secondary programs and institutions (Austin & McCann, 1992). In the 2013 edition of the annual *State of College Admission* published by the National Association for College Admission Counseling (NACAC), the publication reported that high school grades and rigorous courses are the most important factor considered by college admissions officers—more important than standardized tests and class ranking (Clinedinst, Hurley, & Hawkins, 2014).

Organization and guidance. Grades are used by multiple parties for organization and guidance. Counselors often use grades to guide students in taking appropriate high school courses, pursuing college and career paths, and addressing deficiencies (Marzano, 2000). School administrators and counselors use grades “to select, identify, or group students for certain educational paths or programs” (Guskey, 2015, p. 13). Grades are also used to qualify or disqualify

students for various tracks, levels, and special programs. This purpose is similar to the administrative purpose, but in this case the decisions made with grades are less likely to be a part of official school policy, but rather as unofficial policy or made on a case-by-case basis (Marzano, 2000).

Planning and evaluation. Grades are used by educators for planning and evaluation. Teachers often use grades to assess the effectiveness of the teaching of a unit, and schools may use grades as a data source to judge the value and effectiveness of school programs (Guskey, 2015; Guskey & Bailey, 2001; Marzano, 2000).

Motivation. Grades may also be used by multiple parties for motivation. Many believe that grades provide incentives for students to learn, and in this purpose they serve as a type of educational currency. This use is particularly controversial, but nevertheless it is a popular purpose of grading by individual teachers, departments, schools, and districts (Guskey, 2015; Marzano, 2000). Using grades for motivation essentially provides students with punishments and rewards. Students who display desired attributes such as behavior, attitude, and effort are often rewarded with higher grades, while those displaying undesirable attributes are often punished with lower grades (Reeves, 2011).

Researcher recommendations. Although grades are used for a variety of reasons, including the five major purposes explained above, most are not recommended by researchers. Instead, researchers recommend that grades be used for limited purposes and that all interested parties be fully aware of these purposes.

Purpose of grading. In stark contrast to the grading controversy at the high school site level, measurement specialists are in consensus on the purpose of grades (Brookhart, 1999; Cross & Frary, 1999). Brookhart (1999), a leading researcher in the field of measurement, succinctly stated, “Grading is a way to

communicate information about a student's achievement in a course" (p. 69). In other words, Brookhart (1999), along with a consensus of grading specialists, recommends utilizing grades *only* for the communication of student academic achievement. This does not mean that grades cannot be used for secondary purposes such as program evaluation, for example, but the recommendation by researchers is for teachers to create student grades solely to represent student achievement (Brookhart, 1991, 1994, 1999, 2009; Guskey, 2015; Marzano, 2000; Reeves, 2011; Stiggins, et al., 1989). This clear understanding of the purpose of grades is essential for effective use of grades (Guskey & Bailey, 2001). Additionally, Austin and McCann (1992) emphasized that the meaningfulness of grades is dependent upon the community's collective understanding of their meaning.

According to Randall and Engelhard (2010), when teachers follow the recommendations of researchers, grades have the potential to "(a) enable teachers to compare the knowledge and skills of current students, (b) allow teachers to ascertain accurately the preparedness/readiness of incoming students, and (c) provide parents and students with a clear picture of each child's knowledge and understanding of course content" (p. 1376). In other words, if grading is *only* based on student achievement, grades may be used to informally compare students, assess readiness, and clearly communicate student achievement with parents and students (Randall & Engelhard, 2010).

Dimensions of grading. In a seminal study to document teacher grading practices and discrepancies from researcher recommendations, Stiggins et al. (1989) conducted a case study of 15 math, science, social studies, and English teachers in a single high school. The researchers utilized survey, observations, and interviews to compare observed teacher grading practices on 19 specific

dimensions of grading recommended from a sample of introductory measurement textbooks. Stiggins et al. explained that these dimensions did not represent a consensus among researchers; however, they did provide a comprehensive list of recommendations. Below is a synopsis of the 19 researcher recommended practices used in the Stiggins et al.'s study along with support from additional contemporary studies to provide a more detailed description of researcher recommended grading practices.

Communication and grading factors. Stiggins et al. (1989) described the recommendation to communicate grading methods to students. Researchers recommend teachers to clearly explain how they calculate student grades from the beginning of a grading period. Reeves (2011) devoted the first chapter of his book *Elements of Grading: A Guide to Effective Practice* to another form of communication: feedback. In this chapter Reeves explained that grading is feedback, and he went on to illustrate the importance of this feedback and four characteristics of effective feedback: accuracy, fairness, specificity, and timeliness. Reeves' explanations on feedback were largely based upon part of Hattie's (2009) work in *Visible Learning*. Hattie's comprehensive study, which consisted of a meta-analysis of over 800 meta-analyses related to student achievement, ranked feedback as 10 out of 138 different factors for effect size (d), with $d = 0.73$ for effect on student achievement. This effect size is significant, as Hattie explained that when comparing educational outcomes, $d = 0.2$ is small, $d = 0.4$ is medium, and $d = 0.8$ is large. For comparison sake of effect sizes, parental involvement was found to be medium to large ($d = 0.51$), drug use was small to medium ($d = 0.33$), and teacher effect was small to medium ($d = 0.32$). Finally, Waltman and Frisbie (1994) conducted a study on grade interpretation based on the results of the Stiggins et al. study, and they emphasized the importance of

communication between teachers and parents on grade meanings, purposes, and distributions.

As mentioned above, researchers recommend utilizing academic achievement as the single factor in determining grades. Additionally, researchers explicitly state *not* to use learning ability, attitude, motivation and effort, interest, and personality as factors in determining student grades because of the difficulty in objectively defining requirements and assessing these requirements (Allen, 2005; Cross & Frary, 1999; Randall & Engelhard, 2010; Stiggins et al., 1989).

Measuring student achievement. Researchers recommend using written tests, including essays and objective tests, as the primary means of measuring student achievement. However, they also suggest using performance assessments to measure student achievement when it is necessary and done objectively. Stiggins et al. (1989) found that measurement texts state *not* to use either daily written assignments or oral questioning for data on student achievement; however, these sources of data may be used for formative assessment. Researchers recommend gathering enough evidence of student achievement over the course of the grading period to accurately judge the proportion of material mastered by the student; however, too much graded work (as in grading everything) is not ideal either. The recommendation is “typically more than two or three periodic, high-quality assessments are recommended for a quarterly grading pool” (Stiggins et al., 1989, p. 9). Also, when measuring, teachers should ensure a high quality of grading data that is valid, reliable, and not extremely time intensive (Guskey & Bailey, 2001; Stiggins et al., 1989). Messick (1989) posited that grading validity—the extent to which a grade is performing its intended function—is very complex. Messick explained that validity includes not only construct validity, but also value

implications, relevance, usefulness, and social consequences. Finally, Stiggins et al. added that teachers should follow pre-established district grading policies.

Grade calculations. Teachers should use consistent score scales on assignments and apply appropriate pre-determined weighting if desired (Stiggins et al., 1989). Also, teachers should apply a fixed-percentage grade cutoff to letter grades (e.g., A: 90-100, B: 90-89, etc.). Stiggins et al. explained that teachers should *not* apply a normal distribution as the standard for judging grade distributions in single classes or use total point accumulation methods. Finally, when deciding on student borderline grades, teachers should use additional achievement data in order to make an objective decision (Stiggins et al., 1989).

Contemporary grading practices and concerns. Although researchers have found that the main factor that teachers currently use to create student grades is academic achievement (Brookhart, 1994; Randall & Engelhard, 2009), this is not the only factor, and the use of additional factors widely varies. The results of Stiggins et al.'s (1989) seminal case study showed that teachers failed to follow 10 of the 19 recommended dimensions of grading: using only achievement for grades; not using ability, effort, or daily assignments for grades; limiting the amount of data collected for grades; ensuring quality data for grades; aggregating grading components; using fixed percentage grade cutoff scores; refraining from using total point accumulation for grades; and using achievement data for borderline grades. The findings of this often-cited study set the stage for further studies to both document these discrepancies and explore additional problems in teacher grading practices.

Hodgepodge grading. Brookhart (1991) first used the term *hodgepodge grade* to describe current teacher grading practices in an often-cited article entitled "Grading Practices and Validity." Brookhart (1991) criticized current teacher

grading practices by stating “a hodgepodge grade of attitude, effort, and achievement, created in an attempt to provide positive feedback to the student about himself or herself, is not the answer. Such a hodgepodge grade also falls down under a validity check; it does not possess the characteristic of interpretability” (p. 36). This article popularized the phrase *hodgepodge grading* among researchers in the measurement field, and researchers continue to use the phrase to describe teacher grading practices that apply a highly variable patchwork of achievement and non-achievement factors to create a single student grade (Brookhart, 1991; Cross & Frary, 1999; McMillan, 2003).

A study by Cross and Frary (1999) that surveyed 307 middle and high school teachers and 8,664 students in a large Virginia school district on teacher grading practices supported the findings of previous studies (e.g., Brookhart, 1994; Frary et al., 1993) that teachers largely employ hodgepodge grading practices to create and assign student grades. Especially interesting was the finding that most students understood and agreed with these subjective practices. The authors went on to explain that teachers may employ hodgepodge grading to protect both students and themselves from consequences and implications of low grades resulting from creating grades strictly from academic achievement. Despite these findings, McMillan (2001) stated that we must work to fully understand teacher hodgepodge grading practices.

These practices of teacher hodgepodge grading greatly differ from researcher recommendations. In a large study, Frary et al. (1993) surveyed 800 randomly selected teachers to determine the need for measurement training and remediation in schools. Their results showed that teacher beliefs often differed from their own practices. As an example, most teachers agreed that multiple-choice tests are undesirable, yet 52% reported that they frequently or always used

them to assess students. The teachers in this study failed to follow the grading recommendations of researchers, as they indicated that homework, ability, behavior, and improvement all had a part in grading practices. Frary et al. articulated the problem with current grading practices in their finding that there is “widespread disagreement between teacher beliefs and practices and what we and many other measurement specialists would recommend” (p. 28).

Non-achievement factors. Teachers commonly report that they consider many different factors in addition to achievement in determining course grades such as effort, ability, attendance, disruptive behavior, improvement, and assignment completion (Brookhart, 1994; Cross & Frary, 1996; Cizek et al., 1995/1996; Frary et al., 1993; Randall & Engelhard, 2010; Willingham et al., 2002). To further complicate matters, it seems that teachers use non-achievement factors such as those listed above in different proportions and combinations across classes and within their own classrooms (Cizek et al., 1995/1996). Three specific non-achievement factors—effort, behavior, and ability—often play a significant role in teacher grading practices, even though researchers are insistent that this should not occur.

In a study of the grading practices of general education teachers with students both with and without disabilities in the same classroom, Bursuck et al. (1996) found that “while general education teachers, particularly at the junior high and high school levels, are reluctant to pass students no matter what, they do seem somewhat more receptive to passing students who make an effort” (p. 313). The authors warned that grading on student effort inflates grades and can prevent students with and without special needs from receiving the attention they need for improvement, which can lead to further difficulties both in school and the workplace (Bursuck et al., 1996). Brookhart (1994) also found that teachers feel

the need to include effort in the grades of students who work hard, and teachers most often use effort in the grades of students who display low ability. McMillan (2001) recommended that teachers consider the impact of utilizing each grading component, as “teachers who reward effort may be leading students who are not competent, and their parents, to believe that they do in fact demonstrate needed knowledge and skills” (p. 31). McMillan (2001) went on to explain that this practice would be most detrimental for low-achieving students—who may be rewarded with effort to ensure continued engagement in class—if they don’t receive the feedback they need to improve deficient skills.

In a study examining the factors used by teachers to create student grades, Randall and Engelhard (2010) utilized a questionnaire containing 54 different grading scenarios with a sample of 516 K-12 teachers. The authors found that certain grading situations and circumstances, such as borderline grades, resulted in teacher grading practices that were even more subjective than usual. Borderline grades were defined as circumstances in which a student grade is very close to the higher successive grade. For example, an 88.7% is considered a borderline grade, as most teachers consider it a B, but it is very close to a grade of an A, which by many grading scales would begin at 90%. Randall and Engelhard (2010) found several important grading trends: Students on the borderline between an A and B showing low effort and poor behavior often received a B; students displaying low achievement (69% in the provided scenarios) and low ability with excellent behavior and high effort received an average of a 76.8% (C); and overall, regardless of effort and ability, teachers regarded behavior as a very important factor in determining borderline grades. In a research article using the same data set, Randall and Engelhard (2009) found that although student academic

achievement was the primary criterion utilized to assign grades, teachers also used ability, behavior, and effort to grade.

Grades are also misused as a method of punishment and reward, which is part of many teachers' classroom management strategies to control student behavior (Brookhart, 1994). When grades are used for these purposes, students who are rewarded for sought-after behaviors such as homework completion, class participation, and overall good behavior receive inflated grades, and students who display the opposite receive deflated grades. Whether grades are inflated or deflated, neither accurately displays students' academic achievement levels (Hanover Research, 2011). Farkas et al. (1990) found that teachers often use grades to reward student basic skills and work habits. The results of their study of the methods in which teachers rewarded informal standards using data on 486 middle school students showed that "teacher judgments of student non-cognitive characteristics are powerful determinants of course grades, even when student cognitive performance is controlled" (p. 140). Brookhart (1999) stated that although some studies support the positive effects of using grades as rewards and motivation, there is no evidence that supports using grades as negative reinforcement or punishment.

McMillan (2001) conducted a study to determine the relationship between teachers' grading practices and grade level, subject area, and ability level. Using a sample of 1,483 6-12 English, math, social studies, and science teachers from 53 different Virginia schools, the author found only a small relationship between grading practices and grade level. However, McMillan (2001) found a strong positive relationship between student ability level and the use of academic achievement in grading ($d = 0.78$) and a negative relationship between ability level and the use of homework in classroom grades. The findings implied a greater

amount of non-achievement factors were used in grading students of lower achievement levels, thus providing lower feedback on achievement to these students (McMillan, 2001). Because teachers often grade students of high ability tougher than those of average to low ability, students may be at a disadvantage if teachers perceive them to be at a high ability level (Randall & Engelhard, 2009). This practice would seem to encourage students to display lower levels of achievement (at least at the beginning of a school year) and rely on their teachers to reward them for non-achievement factors such as effort and improvement.

Variability in practices. Grading practices in American schools are highly variable. Cizek et al. (1995/1996) declared, “Grades appear to consist of a potpourri of elements that vary from district to district, from teacher to teacher within a district, and even from student to student within a classroom” (p. 174). As grading practices differ among school districts, schools, and teachers (Brookhart, 1994), course grades lose much of their meaning, and lost with this meaning is the potential positive effects of grades on student achievement. In a study of 144 school districts to determine the extent to which districts have similar grading purposes and systems, Austin and McCann (1992) found great inter- and intra-district variation in grading policies and procedures. Of the 71 districts that provided grading documents from multiple levels of districts (department level, school level, and school board/district level), 46 failed to show consistency in grading criteria, while 75 of 90 districts that provided any type of documents on grading policies asked teachers to apply multiple criteria in grading students (Austin & McCann, 1992). In a literature review on 19 grading studies, Brookhart (1994) found a great amount of variation between teachers’ practices with differences in the meaning of grades, purposes, and grading criteria. In their 2010 study, Randall and Engelhard found that teachers differed in their leniency and

severity of grading of the same student. This finding supports the claim that grading by high school teachers is often subjective and highly erratic (Brookhart, 1994; Cross & Frary, 1999).

Zwick and Green (2007) found fairly low variation of high school grade point average (HSGPA) between schools (15.83%), while the variation between schools of SAT math (26.68%) and SAT verbal (26.06) scores was much higher. Thus, regardless of student true academic achievement (displayed by SAT scores), all schools seem to assign the same frequency and range of student grades. Additionally, it should be noted that variance greatly differed by ethnic group, as it was lowest for White students and highest for Asian American students (Zwick & Green, 2007).

Communication. Parents generally assume that grades represent their child's achievement in a class (Brookhart, 1994); however, as grades are often inaccurate representations of student achievement, parents may be misinformed of their child's true performance in their classes. Waltman and Frisbie (1994) conducted a study with a sample of 16 different fourth-grade math teachers and 246 parents from 16 different Iowa schools. Results showed that teacher-parent communication was muddled, and they found that teachers and parents showed a great amount of inconsistency and variability in the interpretation of student grade purposes, meanings, and distributions. Randall and Engelhard (2010) pointed out that while American students score low on state, national, and international standardized tests, teacher assigned course grades do not reflect this low level of achievement. This would lead many to question the meaning of course grades in the United States. If grades do not represent student achievement, then for what future purpose do they serve? Because it is well established that grades are often based upon measures other than student achievement (see the hodgepodge grading

section above), the academic community largely considers course grades to be a subjective measure of student learning (Guskey, 2015). This is so much of an issue that standardized test scores are often used instead of grades when many interested parties need to make a major decision regarding a student (Randall & Engelhard, 2010).

Validity. Researchers make it clear that current grading practices often lack validity (Allen, 2005; Brookhart, 1991, 1994). Allen (2005) explained that grading validity includes the extent to which grades reflect and communicate student academic achievement. Allen added that current invalid grading begets future invalid practices, as teachers often replicate the invalid grading practices that they experienced as students in the past partly because current grading preparation and training is often inadequate to override the influence of the past experiences (Allen, 2005). Brookhart (1991) blamed current discrepancies in grading practices on validity problems. However, unlike many researchers, the author explained that teachers often consider factors other than construct validity in grading because of the contextual realities with which they face. Brookhart (1991) concluded that teachers often inadvertently apply Messick's (1989) widely accepted explanation that grading validity must also consider value implications, relevance and usefulness, and social consequences. This explanation implies that teacher grading validity is actually greater than many researchers credit because of the contextual considerations.

Grades versus test scores. Brennan et al. (2001) studied differences in equity between teacher grades and high-stakes tests. The authors analyzed teacher assigned grades and Massachusetts Comprehensive Assessment System (MCAS) scores of 736 eighth-grade students from six Boston middle schools. Results showed that "MCAS hurts the average competitive position of African American

students in math and of girls in math and science” (Brennan et al., 2001, p. 206) as compared to teacher assigned grades. The authors explained these differences were likely due to the highly subjective nature of teacher grades, including factors such as behavior, attitude, and effort, which teachers often used to compensate for low student achievement. Brennan et al. (2001) concluded that grades are usually more equitable than standardized tests, yet they are less accurate measures of achievement. The authors recommended using a combination of standardized test scores, student grades, and perhaps other measures of student academic and non-academic achievement to make educational decisions.

Goldwater and Nutt (1999) conducted a study of the grading practices of 20 teachers and 101 of their students to determine the relationship between the compatibility of student and teacher backgrounds and teacher grading practices. To determine this relationship, the authors used a subjective grading index, in which the subjectivity of teacher assigned grades was determined by comparing these grades to student final exam grades. Results showed that students who displayed similar family backgrounds to their teachers received higher and more subjective grades than students with different family backgrounds (Goldwater & Nutt, 1999).

Alternative perspectives. Despite the documentation of problematic grading, educators do display positive attributes in their practices. In fact, some researchers, such as Bowers (2009, 2011), have stated that the research community must stop chastising teachers for failing to follow grading recommendations. Bowers (2009, 2011) asserted that researchers must not only acknowledge the complexities in grading, but also consider teacher-created grades as valuable assessments of multiple attributes of student success. However, even these positives are often tempered with exceptions. It is true that studies consistently

show that academic achievement is the primary factor used by teachers in determining student grades at the high school level (Randall & Engelhard, 2009), and teachers predominantly measure these achievement levels through the use of quizzes and tests (Brookhart, 1994; Frary et al., 1993). However, Brookhart (1994) pointed out that some teachers merely use achievement data from tests to confirm pre-established judgments of student achievement. As previously mentioned, teachers agree that in theory, grades represent student achievement in a particular class; however, discrepancies exist between teacher grading philosophy and grading practices (Randall & Engelhard, 2009). Many teachers realize that there are problems with their own practices, but it seems that a lack of viable alternatives prevents them from making significant changes that might better align their practice with their philosophy (Guskey, 2001).

But teachers do not ignore all recommendations of measurement specialists; they do attempt to follow the recommendation of researchers to clearly communicate their grading policy to students and parents (Brookhart, 1994). Additionally, fairness is an important issue to teachers in the practice of grading (Brookhart, 1993, 1994; Zoekler, 2007). In fact, studies show that a major reason that teachers do not base grades completely on achievement is because they feel the practice would lack a sense of equity (Frary et al., 1993). To ensure this equity, Frary et al. (1993) found that teachers reward students for hard work and engagement in their classes by utilizing homework and participation as a common component of grades. In a study on teacher assessment practices that used a sample of 143 Midwestern elementary and secondary teachers in a master's level course on measurement and evaluation, Cizek et al. (1995/1996) reported that a commonality of all the teachers was a "success bias," which they explained as teachers compiling information from tests and quizzes in ways that would result in

the highest possible grade for each student. Teachers really want students to succeed, and they are willing to use grading practices to ensure that they receive good grades (McMillan, 2001). McMillan (2003) referred to this as *pulling for students*, and he found this to be a common grading practice of teachers. Clearly, teachers deeply care for their students, but grading practices such as those described in the lines above create some significant problems in education.

Even with the widely established discrepancy between recommendations and practice, some researchers believe that grades are effective in performing multiple purposes. In *What's In a Grade? The Multidimensional Nature of What Teacher-Assigned Grades Assess in High School*, Bowers (2011) displayed his belief in the importance of teacher assigned grades by stating, "Recently, this dualistic nature of grades has been explored as useful data as a multidimensional assessment that assesses both academic knowledge and non-academic behaviors" (p. 143). Bowers (2011) explained that grades might be an effective measurement of both academic achievement and non-cognitive social behaviors that are important for educational success. In a 2008 study, Bowers found that about 25% of grading could be attributed to academic achievement, while about 75% is made of social processes. While stating further research needs to be done to better understand this 75%, Bowers (2009) claimed that this portion of the grade is evidence of a *success at school factor* (SSF), which has a tight connection to academic knowledge. Bowers added that as teacher assigned grades communicate both academic achievement and social processes, remediation efforts for students with low grades should address both academic and social deficiencies.

In a study using 1988 NELS data from 8,454 high school seniors, Willingham et al. (2002) developed a framework to explain the difference between standardized test scores and teacher-assigned grades. This framework found

student grades to show a strong relationship with standardized test scores after controlling for five factors: subjects covered, grading variation, reliability, student characteristics, and teacher ratings. Upon accounting for the five factors, Willingham et al. (2002) were able to increase the correlation of student grades and test scores from 0.62 to 0.90. Grading variation and teacher ratings, along with a principle they referred to as *scholastic engagement*, were shown to be major factors in the framework. Upon reporting these results, the authors stated that the fact that these three major factors could be easily accounted for in the group and individual differences shows the intrinsic validity and fairness in tests and grades (Willingham et al., 2002).

Teacher Grading Decision Making

While the studies above document the discrepancy between researcher recommendations and teacher practices, we must better understand the reasons for these practices to move beyond simply reporting and describing teacher hodgepodge grading (McMillan & Lawson, 2001). Randall and Engelhard (2010) claimed that teachers rely on a combination of factors such as personal philosophy, college classes and professional development, school or district policy, and perceived consequences to create course grades for their students, while Allen (2005) explained that the literature on grading practices shows that educators do not make grading decisions that are based on principles of validity. A better understanding of the often-puzzling process that teachers utilize to create grades may be achieved through a study of the teacher grading decision-making process. Because a number of researchers agree that effective grading requires an effective decision-making process (Cauley & McMillan, 2000; McMillan, 2003), an analysis of the components of the process may lead to a better understanding of the current process and possible solutions. This section of the literature review

presents a description of the studies that led to McMillan's (2003) teacher grading decision-making framework (originally shown and described in Chapter 1) and a review of literature on major components of the framework.

Teacher Grading Decision-Making Process

The teacher grading decision-making process was used as one of the frameworks in this study to compare the grading process in schools of different poverty levels. McMillan and Nash (2000) explained, "Because of the interplay between the teachers' beliefs, external factors, and student/classroom characteristics, a great amount of variety in classroom assessment and grading is evident" (p. 31). The framework began with McMillan and Nash's (2000) qualitative study on the assessment and grading decision-making practices of elementary and secondary math and English teachers, which was presented as a report at the Annual Meeting of the National Council on Measurement in Education. The researchers interviewed 24 elementary and secondary math and English teachers from 13 schools and 7 districts. Analysis of interview data produced six emergent themes in assessment and grading decision making: teacher beliefs and values, classroom realities, external factors, teacher decision-making rationale, assessment practices, and grading practices. McMillan and Nash (2000) used these themes to create a preliminary teacher assessment and grading decision-making model. This model described the resulting tension in the process that results from the interaction between teacher beliefs and values, external realities, and external factors. McMillan and Nash (2000) explained that this interaction of factors led to teacher decision-making rationale, which resulted in assessment and grading practices. The authors stated, "The main tenet of the model holds that that there was tension between the internal beliefs and values of

teachers and the realities of their classroom environments and other external factors that are imposed on them” (McMillan & Nash, 2000, p. 6).

McMillan (2003) expanded upon this original study in a follow-up article to further explore assessment and grading decision making and create a thorough theoretical framework. This paper infused concepts and ideas from relevant literature with data from the original study to bolster the original teacher assessment and grading decision-making model into a more comprehensive framework. The resulting decision-making framework, which is one of the two theoretical frameworks of this dissertation (presented in Chapter 1, Figure 1, p. 13), contains three major domains that explain teacher grading decision making: influences of grading (i.e., teacher knowledge, beliefs, expectations, and values; external factors; and classroom realities), decision-making rationale, and grading practices (McMillan, 2003). The framework was intended to explain both assessment and grading decision making, but as this study focuses only on grading, only aspects of grading decision making are considered.

Teacher Preparation and Training

Many teachers are underprepared and under trained to successfully create student grades congruent with researcher recommendations. Stiggins et al. (1989) suggested that a possible reason for the discrepancy between researcher recommendations and actual teacher practices is either because teachers are largely unaware of the research on grading or because they are not trained to employ the suggested methods. Teachers continue to report that they receive little to no education in their teacher preparation classes on school measurement and grading (Brookhart, 1994; Randall & Engelhard, 2010), and they often fail to receive professional development on the grading while in the classroom (Frery et al., 1993). Despite the fact that teachers often report a lack of knowledge and skill

grading in the classroom, their needs seem to continue to go unfulfilled. Hills (1991) suggested that the lack of training in grading results not only in an ignorance in the practice, but also in an overall sense of apathy on the topic. However, there are also teachers that take measurement courses and receive training, yet fail to use successful grading practices (Hills, 1991).

Airasian and Jones (1993) explained that a reason measurement courses have such little impact on teacher grading practices is because of the disconnect between college courses and the reality of the classroom environment. The authors explained that measurement courses often present only best grading practices, but teachers frequently feel they are unrealistic given the demands they face in the classroom. Additional studies have shown that teacher education and training in measurement and grading have little effect on the objectivity of teacher grades (e.g., Brookhart, 1991, 1993). The vast majority of teachers that do receive schooling in measurement during teacher preparation classes or training in methods of grading while teaching do not incorporate these ideas learned into their grading practices (Brookhart, 1994; Randall & Engelhard, 2010). Brookhart (1993) attributed this surface level impact on the fact that typical grading and measurement training completed by teachers is similar to introductory measurement courses, which primarily focus on grading philosophy.

Frary et al. (1993) suggested that current training in measurement and grading is insufficient, and that future classes should address current controversies in grading and focus on the main deficiencies. However, little study has been done on the small group of teachers that do grade by the recommendations of researchers after taking courses or completing training. This could potentially lead to strategies of bridging the gap between theory and practice. Brookhart (1993) did find that although training in grading and measurement showed no impact on

grading practice, it did affect the grading philosophy of teachers. For example, ability and improvement were significant considerations in the grading philosophies of teachers without measurement training; however, these non-achievement factors were not considerations in the grading philosophy of teachers with measurement training. Randall and Engelhard (2010) called on teacher preparation programs and school administrators to utilize the results from grading studies to address grading deficiencies in teacher education classes and professional development.

Teacher Feelings and Roles

Despite the numerous studies citing the faulty grading practices of high school teachers, few researchers doubt their intentions. Teachers care about their students—this is the primary reason they became teachers in the first place. But with such emotional connections to their students, it may be difficult for teachers to withhold this urge to help when it comes time to grading (Cziek et al., 1995/1996). These feelings may mask the effectiveness and intended purpose of course grades, as empathy and encouragement often make their way into the grading practices of many teachers (Cross & Frary, 1999). On a philosophical level, most teachers, just like researchers, agree that grades should represent student achievement (Cross & Frary, 1999). However, Brookhart (1993) reported that when it comes to the practice of grading (as opposed to the philosophy), teachers possess a belief that students must earn their grades, and they feel that by assigning course grades, teachers are rewarding their students for their work completed.

With all of the issues in grading, it should be no surprise that teachers often dislike the practice, as they often view it as a difficult one (Barnes, 1985). Thorndike (2005) referred to grading as both an unpleasant and anxiety-provoking

experience for teachers. Although teachers may have a difficult time articulating the source of their discomfort and anxiety when it comes to grading, Brookhart (1994) explained that it is a direct result of having to play both the role of the advocate and the judge. Brookhart (1994) added that teachers play the role of the judge in determining students' level of achievement, but they also continually play the role of the advocate by advocating for the students' success. Brookhart (1993) stated that teachers feel uncomfortable with the "conflict between the recommended practices, which concentrate on grade interpretation and meaning, and concerns with the uses of grades" (p. 139). Brookhart (1993) went on to state that teachers have strong feelings for the values and social consequences of grading, which is related to their role as student advocates. This role seems to be closely related to teachers caring for their students and the reason many went into the profession in the first place.

Bishop (1992) emphasized the incompatibility of teachers playing roles of both the student advocate and judge of achievement. These discrepancies between philosophy and practice display the high level of cognitive dissonance experienced by teachers because of the multiple roles they are forced to play in the practice of grading. Bishop (1992) suggested that solving this discrepancy between the two conflicting roles requires teachers to focus on the advocate role and delegate the judge role to an outside party. However, teachers are often placed in very difficult situations by both researchers and school administrators, as they are often required to solve the discrepancy of the conflicting roles on their own (Brookhart, 1993). Airasian and Jones (1993) further explained sources of grading tension by describing teacher grading roles requiring both social and technical expertise. Teachers are never taught or directed how to merge both roles in grading practice; instead, university measurement classes teach the technical requirements, and

teachers are left to themselves to discover the social component (Airasian & Jones, 1993).

External Pressure

Both Cross and Frary (1999) and Iacus and Poro (2011) suggested that teachers make the choice to grade against the recommendations of researchers because of external pressure from other teachers, parents, students, and administrators. Teachers often feel pressure to submit grades that appear similar to that of a normal distribution. Because grading strictly based on student achievement often lowers student grades, teachers can increase the number of upper level grades and decrease the number of failures by including factors such as effort, participation, and behavior into students' final grades. However, this idea does not explain the subjective grading by teachers who do not experience significant pressure from the aforementioned parties (Cross & Frary, 1999; Iacus & Poro, 2011).

In their study of the grading practices of 516 teachers from the same district with a uniform district grading policy, Randall and Engelhard (2010) found that teachers mostly followed the district directive to grade only based on student achievement. However, teachers reported that in situations such as borderline grading, teachers often used factors such as behavior and effort to make the final decision, despite the district policy.

Agnew (1985) conducted a study of teacher grading practices of 277 teachers from seven schools in the San Francisco Bay area, and the author found that teachers in a low SES school not only used a greater amount of non-achievement factors to determine grades, but they also displayed the greatest amount of dissatisfaction with grading standards and their own grading practices. This high dissatisfaction rate implied a high amount of tension in the teacher

decision-making process. Agnew found that at least part of this dissatisfaction came from the impact of the unwritten rule that teachers in their study maintained grading autonomy as long as they did not fail anyone. This pressure from administration made an obvious impact on the manner in which teachers conducted their grading practices (Agnew, 1985).

Classroom Context

Teachers often feel that grading practices must address certain classroom contextual realities. Teachers feel the need to grade on factors such as effort, behavior, and homework completion in order to ensure discipline and a sense of power in the classroom (Cross & Frary, 1999). Teachers often feel that if homework is not graded, then students will not complete it. They also often feel the need to use grades in a punitive manner as part of a classroom management strategy (Hills, 1991).

Motivation is a consideration in the grading practices of many teachers. Docan (1996) conducted a mixed-methods study to examine the ways in which grading systems motivate students. Using a sample of 101 university students, Docan studied the effects of two different grading systems: one that began students with zero points and allowed them to accrue points to earn a grade (students referred to as “earners”), and one that started students with maximum points and required students to maintain their grade throughout the semester (students referred to as “maintainers”). Quantitative results showed that maintainers displayed slightly higher levels of motivation. Qualitative results showed that maintainers were motivated by satisfaction in the system, unfamiliarity with the system, stress, and punishment, while earners were motivated by familiarity with the system and rewards. The author concluded that

grading systems may impact the ways in which students perceive the subject matter, and further types of motivation should be investigated (Docan, 1996).

As an alternative method of addressing student motivational issues, some schools have turned to minimum grading as a solution. Guskey (2015) explained, “school districts implement these minimum-grade policies simply to eliminate the confounding effects of a zero in percentage grading systems” (p. 31). Minimum grading policies allocate a minimum grade to individual student scores (often a 50%) in an effort to rectify the inherent bias toward failing scores in the 100-point grading scale. In a study of 7 years of data in an urban high school that used a minimum grading policy, Carey and Carifio (2012) found that minimum grading did not promote grade inflation or social promotion, while it did correspond with a decrease in the school’s attrition rate during this time from 16.1% to 13.5% and a savings of over \$1,000,000 in reduced summer school costs.

The Impact of Poverty on Teacher Grading

This final section of the literature review is closest to the topic of the current study. The literature on poverty and teacher grading practices is sparse, yet significant, as it addresses a topic that affects children of high-needs. The handful of studies reviewed below sets the stage for the three chapters that follow. These studies address differences across school poverty levels, differential grading, and predictions.

Differences across School Poverty Levels

Schools of high-poverty levels have been shown to utilize different grading practices than schools of low-poverty levels. In Agnew’s (1985) study of the grading practices of 300 teachers from seven high schools in and around San Francisco, the author found that teachers in a school with the lowest SES level and

the most minority students least utilized student learning in grading. Instead, teachers heavily relied on effort, behavior, and attendance to formulate student grades. In discussing the implications of the findings, the author remarked, “Low status and minority students are graded not on their academic achievements, but on their attendance and deportment. The value questions this practice raises are self-evident and worthy of public discussion” (Agnew, 1985, pp. 41-42).

McMillan et al. (2002) cautioned educators about the use of effort as a reward for all students, but low-SES students in particular. Additionally in an often-cited report on the meaning of student grades, the Office of Educational Research and Improvement (1994) analyzed national eighth grade student data from the NELS of 1988. The study found that student grades in high-poverty schools (defined as schools in which more than 75% of students received FRPL) appeared inflated and did not accurately provide feedback on student achievement. The study further reported that students in high-poverty schools who received mostly A’s in English class earned about the same standardized test scores in reading as students who received mostly C’s and D’s in low-poverty schools. In math, students in high-poverty schools who received mostly A’s earned about the same standardized test scores as students who received mostly D’s in low-poverty schools (Office of Educational Research and Improvement, 1994).

Cauley and McMillan (2000) surveyed 633 middle school teachers to determine the difference in grading practices between low- and high-poverty schools. While the study results showed no difference in student ability levels, high-poverty schools were more likely to consider non-achievement factors such as disruptive student behavior, attention and participation, effort, improvement, work habits, neatness, and extra credit. At the same time, low-poverty schools were more likely to create grades mostly or completely from student academic

achievement. The authors concluded that teachers in high-poverty schools “who reward effort and other non-academic factors may allow students who are not competent, and their parents, to believe that they in fact demonstrate needed knowledge and skills. This may, in turn, lead to a distorted perception of ability” (Cauley & McMillan, 2000, p. 11).

Differential Grading

Hanna and Linden (2012) investigated discrimination in grading practices in schools in India. The study utilized 120 teachers to grade assessments of 69 different students. All assessments were blindly graded except for the randomly assigned student demographic information on the cover of each exam. Results showed that teachers displayed a small amount of discrimination against exams believed to be completed by students of low-castes, as higher-caste exams received scores 0.03-0.08 standard deviations higher than those of lower castes (1.5 percentage points). The authors concluded that teachers discriminated when they were uncertain about the grading instrument (Hana & Linden, 2012). Madon et al. (1998) studied teacher perceptions and stereotypes in the classroom with a sample of 49 to 56 seventh-grade math teachers and 2,000 of their students. Their findings showed that teachers perceived high-class students as higher performers, using slightly more effort, and having more talent than lower-class students. Researchers compared teacher perceptions to student data and found teachers to be mostly accurate about student performance and effort. However, results showed that teachers overestimated the differences in talent between students of low and higher classes. Finally, the researchers found that student personal characteristics had a bigger impact on teacher perceptions than assumed social group characteristics (Madon et al., 1998).

Rauschenberg (2014) used 3 years of statewide public high school data from Algebra 1 and English 1 courses in North Carolina to determine differences in course grades. Results showed that student characteristics were indicators of differential grading. Female, limited proficient English, and twelfth grade students earned significantly higher grades when test scores and student, teacher, school, and district characteristics were held constant, while low-income students earned lower grades than those of higher income levels (Rauschenberg, 2014).

Predictions

Guskey (2011) analyzed the records of 8,000 students from five high schools to determine the stability of high school student grades throughout a course. The author found that high school grades are mostly consistent throughout a course, and the first assigned grade is a good predictor of the final course grade. Further, the author found “modest but statistically significant influence of gender, grade level, ethnicity, and poverty level” (Guskey, 2011, p. 95). Poverty level, in particular, was shown to have a negative correlation with first term grades, final grades, and grade changes. Guskey (2011) concluded by stating the relationship between poverty level and both grading stability and grading in general must be further explored.

High school GPA and SAT scores are often used to predict student college success (J. Rothstein, 2004). Zwick and Himelfarb (2011) studied the effect of student SES on the accuracy of student GPA and SAT scores in predicting their first-year college grade point average (FGPA). To do so, the authors analyzed data obtained from the College Board from a total of 123,385 students. Results from the analysis supported their hypothesis that the college success of African American and Latino students (based on SAT and high school GPA) was over predicted because these two groups of students are more likely to attend low-SES

schools, and grades of these schools less accurately communicate student academic achievement. Zwick and Himelfarb concluded by recommending that colleges consider student SES when accepting students for admittance, as predictions of student college success were shown to be more accurate after controlling for SES.

Summary

Schools throughout the United States continue to show an income achievement gap, and schools of low poverty consistently outperform those of mid and high poverty (Borg et al., 2012; Guskey, 2011; Heckman, 2006; Reardon, 2011; Sirin, 2005). Regardless of the poverty level, grading is a significant practice in schools, and grades play a major role in the education and future lives of American students (Guskey, 2015). But while grading practices have more than 350 years of history in the United States and colonial America, for a majority of the time, grading has been controversial and problematic (Brookhart, 2009; Cureton, 1971; Kirschenbaum et al., 1971; Smallwood, 1935). Many of the same problems that we encounter today in grading have been discussed and debated in K-12 schools and universities for over 100 years (Allen, 2005).

A discrepancy between the grading recommendations made by measurement specialists and the grading practices of teachers has existed for over 50 years, yet we continue to struggle with ways in which to solve the problem (Brookhart, 2009; Kirschenbaum et al., 1971). Today, most teachers agree that in theory, grades should only represent student academic achievement; however, in practice, they often incorporate multiple non-achievement factors such as effort, behavior, and ability (in addition to achievement) into grades (Frary et al., 1993). As a result of these subjective grading practices, grades serve few of the intended

purposes (Bowers, 2011; Cizek et al., 1995/1996; Guskey, 2015; Marzano, 2000; McMillan, 2001; Reeves, 2011; Stiggins et al., 1989).

To better understand teacher grading practices and to work toward a solution to the problem, McMillan (2003) created a teacher grading decision-making model that explains three major domains that impact the grading process: influences of grading, teacher grading rationale, and teacher grading practices. This framework may be used to understand reasons for the discrepancy between researcher grading recommendations and teacher grading practices so that more effective recommendations can be made (McMillan, 2003).

Lamentably, grading has been shown to be more subjective and less accurate in high-poverty schools, as teachers often utilize a greater number of subjective factors to determine student grades (Agnew, 1985; Cauley & McMillan, 2000; Office of Educational Research and Improvement, 1994). Conversely in low-poverty schools, teachers are more likely to grade according to the recommendations of researchers, which is to create grades only from student academic achievement (Cauley & McMillan, 2000).

The current study that is presented in the three chapters that follow is built upon the research presented in this chapter. With the aid of a critical pedagogy theoretical framework, the study aimed to add to the understanding of the income achievement gap by analyzing the impact of school poverty level on teacher grading. The study utilized McMillan's (2003) teacher grading decision-making framework to better understand teacher grading practices that have so long differed from researcher recommendations to work towards improvement. This study adds to the literature by combing three variables for what the researcher believes to be the first time: poverty, teacher grading practices, and teacher decision making.

CHAPTER 3: METHODOLOGY

This chapter presents the methodology utilized in the study. The chapter begins with a statement of the purpose, and it is followed by a presentation of the research questions. Next is a detailed description of the research design and the study sample and participants, and this is followed by an explanation of data collection and the pilot study. The chapter concludes with an explanation of the basis of comparison, a description of data analysis, and finally the study limitations.

Purpose of the Study

The purpose of this study was to investigate the effects of school poverty level on teacher grading decision making. A transformative explanatory sequential mixed methods design (Creswell, 2014) was used to address this purpose, and a critical pedagogy theoretical framework was employed as the overarching lens of the study. This framework was used to provide a means for focusing on and analyzing poverty's impact in the study. The mixed methods involved a sequential gathering of quantitative and qualitative data. In the first phase, quantitative data were collected with a survey structured around McMillan's (2003) teacher grading decision-making framework. This was the second theoretical framework used in the study, and it was utilized to collect data from teachers in 17 high schools throughout Sunnyside School District (SSD) (pseudonym) to assess the effect of school poverty level on two domains of the framework: teacher grading practices and the influences of teacher grading. In the second phase, qualitative data were collected both in the previously mentioned survey and in teacher focus groups. The qualitative data were analyzed with a constant comparative analysis method to triangulate the quantitative results on two

domains of the framework and to explore teacher grading rationale at each school poverty level in the third domain of the framework.

Research Questions

The research questions of this study align to McMillan's (2003) teacher grading decision-making framework, as shown in Table 2. This study utilized a single guiding research question: How does school poverty level affect the teacher grading decision-making process?

This guiding question was broken into three specific research questions to address each of the three domains of the teacher grading decision-making framework:

1. How does school poverty level affect teacher grading practices?
2. How does school poverty level affect the influences of teacher grading?
3. What rationale do teachers of different school poverty levels use to make grading decisions?

As displayed in Table 2, the first research question addressed Domain 3 of the teacher grading decision-making framework, and the critical pedagogy lens aided the researcher in analyzing and comparing the objectivity of teacher grading practices across low-, mid-, and high-poverty schools.

The second research question addressed Domain 1 of the decision-making framework, and the critical pedagogy framework helped to analyze and compare the various influences of teacher grading in each school poverty level. Finally, the third research question addressed Domain 2, and the critical pedagogy framework helped to analyze and compare the type and complexity of grading rationale of teachers of each school poverty level.

Table 2

Alignment of Research Questions with Theoretical Frameworks

| Research Question | Teacher Grading Decision-Making Framework Domain | Critical Pedagogy Framework Perspective |
|--|--|--|
| 1) How does school poverty level affect teacher grading practices? | Domain 3: Grading Practices | Objectivity of grading practices by school poverty level |
| 2) How does school poverty level affect the influences of teacher grading? | Domain 1: Influences of Grading (i.e., teacher knowledge, beliefs, expectation, and values; external factors; classroom realities) | Type and impact of influences of grading by school poverty level |
| 3) What rationale do teachers of different poverty levels use to make grading decisions? | Domain 2: Rationale | Type and complexity of rationale by school poverty level |

Research Design

This study utilized a transformative explanatory sequential mixed methods design (Creswell, 2014) to address its research questions. Creswell (2014) explained, “Transformative mixed methods is a design that uses a theoretical lens drawn from social justice or power as an overarching perspective within a design that contains both quantitative and qualitative data” (p. 16). As described in Chapter 1, the study was designed upon a critical pedagogy theoretical framework to address the effect of school poverty on teacher grading decision-making. Within the context of critical pedagogy, McMillan’s (2003) teacher grading decision-making framework was applied to the structure of the study. The methods and research questions of the study align to this framework.

Mixed methods is a relatively new methodology in research (Creswell, 2014). Creswell and Plano Clark (2011) provided a comprehensive definition of

mixed methods research that includes a combination of philosophy, methods, and research design orientation. This definition explained that the mixed methods researcher

- collects and analyzes persuasively and rigorously both qualitative and quantitative data (based on research questions);
- mixes (or integrates or links) the two forms of data concurrently by combining them (or merging them), sequentially by having one build on the other, or embedding one within the other;
- gives priority to one or to both forms of data (in terms of what the research emphasizes);
- uses these procedures in a single study or in multiple phases of a program of study;
- frames these procedures within philosophical worldviews and theoretical lenses; and
- combines procedures into specific research designs that direct the plan for conducting the study. (Creswell & Plano Clark, 2011, p. 5)

This study employed a pragmatist paradigm, as knowledge claims are problem-centered, and the researcher used any approach necessary to address the problem (Creswell, 2003). Creswell (2003) explained, “Pragmatism is not committed to any one system of philosophy and reality. This applies to mixed methods research in that inquirers draw liberally from both quantitative and qualitative assumptions when they engage in their research” (p.12). In the spirit of this statement, a mixed methods approach was chosen because both quantitative and qualitative methods were determined to best address the purpose of the study and collect data on the three domains of the teacher grading decision-making framework.

This research design was fixed in nature, as the methods were determined prior to data collection (Bryman, 2006). The study used three of Bryman’s 16 reasons to utilize mixed methods in a study: addressing research questions that require either quantitative or qualitative methods (Table 3), triangulating quantitative and qualitative data, and increasing the credibility of the study.

Additionally, as this study combined concepts from McMillan and Nash's (2000) qualitative study of teacher grading decision making and Cauley and McMillan's (2000) quantitative study of poverty and grading practices, a mixed-methods approach was viewed as most appropriate.

Table 3

Alignment of Research Questions, Methods, and Instruments

| Research Question | Method | Instrument |
|---|----------------------------|---|
| 1) How does school poverty level affect teacher grading practices? | Quantitative & Qualitative | Survey Questionnaire & Interview Protocol |
| 2) How does school poverty level affect the influences of teacher grading? | Quantitative & Qualitative | Survey Questionnaire & Interview Protocol |
| 3) What rationale do teachers of different poverty levels use to explain their grading practices? | Qualitative | Survey Questionnaire & Interview Protocol |

The independent variable of the study was school poverty level. The dependent variables were teacher grading practices, influences of teacher grading (i.e., teacher knowledge, beliefs, expectations, and values; external factors; and classroom realities), and teacher grading rationale. To collect data on teacher grading practices by school poverty level (Research Question 1), the study primarily employed quantitative methods, although qualitative data were used for triangulation. Data on the three types of influences of grading (Research Question 2) were collected equally through quantitative and qualitative methods, and qualitative methods were also used for triangulation. Data on teacher grading rationale (Research Question 3) were collected through qualitative methods, as this data were sought for descriptive purposes (Table 3).

In a description of the mixed methods design, Bryman (2006) described four key decisions for researchers: “(1) the level of interaction between strands, (2) the relative priority of the strands, (3) the timing of the strands, and (4) the procedures for mixing the strands” (p. 64). In this mixed methods study, quantitative and qualitative strands primarily displayed an *interactive* level of interaction, as methods were mixed both before and during final data interpretation (Bryman, 2006) (Figure 2). More specifically, this interaction of strands occurred in data collection for Research Questions 1, 2, and 3, data interpretation of Research Questions 1 and 2, and in the final interpretation of all collected data to address the overarching research question.

Creswell (2014) explained that explanatory sequential mixed methods is a study that begins with quantitative research in an initial phase and then follows with a qualitative phase to further explain the quantitative results. Because this study was designed for comparative purposes, it prioritized the quantitative research. Qualitative methods were used in a secondary role to triangulate data on Research Questions 1 and 2 and also as the main source of data to address Research Question 3. The design utilized multiphase combination timing, as the study progressed in two distinct phases (Bryman, 2006). Quantitative data were collected via survey questionnaire in Phase 1, and by the end of this phase, grading practices and influences of grading (Survey Parts 2 and 3) were analyzed by school poverty level (and also by school, teaching experience, teaching subject, and class level) (Survey Parts 1 and 2) to address Research Questions 1 and 2 (Figure 2). Results of data collected and analyzed were used to verify comparison groups for Phase 2 and also to refine focus group questions. In Phase 2, focus groups were conducted in each of the school poverty level groups, and upon their completion, qualitative data were analyzed to address Research Question 3 and to

triangulate data in addressing Research Questions 1 and 2 (Figure 2). Results for each of the three research questions were finally aggregated and reinterpreted to address the guiding research question.

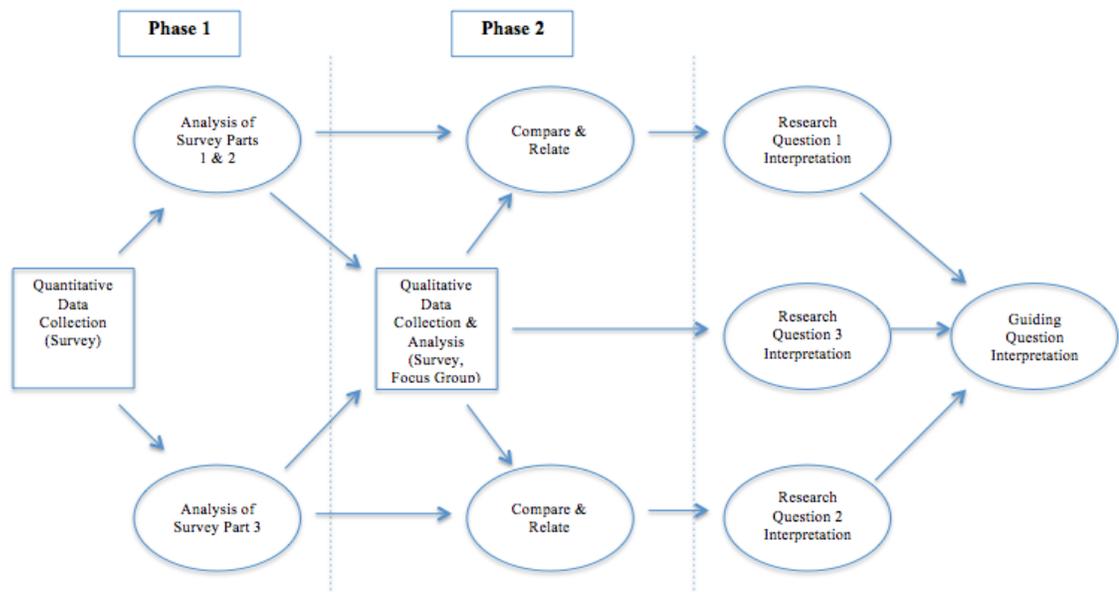


Figure 2. Mixed methods study design.

Quantitative Data

The quantitative component of this overarching mixed-method design utilized a causal-comparative method to investigate the effects of school poverty level on teacher grading practices and influences of teacher grading. Best and Kahn (2006) explained, “Causal-comparative research scrutinizes the relationship among variables in the studies in which the independent variable has already occurred, thus making the study descriptive rather than experimental in nature” (p. 134). This study utilized a survey in the form of a self-administered Web-based questionnaire to collect cross-sectional quantitative data (Appendix A). The use of survey provides a relatively easy, cost-effective method that rapidly produces data available for further analysis (Best & Kahn, 2006; Creswell, 2003). The survey in

this study was utilized to obtain specific information on the teacher grading decision-making process from a relatively large sample in order to infer the application of the process to larger populations (Creswell, 2003). It was selected as an instrument to efficiently determine the grading practices of teachers by school poverty level to address Research Question 1, and it was also used to determine influences of grading by school poverty level to address Research Question 2 (Table 3, p. 91).

Manzo and Burke (2012) explained that one drawback of the survey is the threat to validity caused by low response rates. To address this potential threat, this study utilized multiple steps recommended by the authors to maximize response rates including sending pre-notification emails before sending out the surveys, sending out survey invitations that were clear in purpose and visually friendly, including approximate time for completion, sending between two to four reminder emails, and addressing possible confidentiality concerns (Manzo & Burke, 2012).

Qualitative Data

Qualitative data were collected and analyzed utilizing constant comparative analysis (Corbin & Strauss, 2008; Glaser, 1965). This method was selected because the researcher aimed to gain an understanding of the teacher experience of grading decision making, and constant comparative analysis allowed for a comparison of participants within and across each school poverty level group for this phenomenon (Creswell, 2003). Focus groups were selected as the primary method of qualitative research for multiple reasons: to provide an efficient means of gathering qualitative data from a group of teachers from selected school sites, to allow for observations of teacher interactions on the provided questions, and to allow for teacher interactions to facilitate participant idea generation (Morgan,

1997). In their 2000 study on teacher grading practices, McMillan and Nash interviewed 24 teachers and found many unable to describe the rationale for their practices. They stated their goal was to “‘get inside the head’ of teachers” (McMillan & Nash, 2000, p. 36), but utilizing one-on-one interviews as a method of data collection may have limited teachers’ ability to describe their rationale, thereby limiting results. Focus groups were specifically used in this study to address this potential limitation. Morgan (1997) supported this idea by explaining that focus groups may exceed interviews in depth of information “for topics that are either habit-ridden or not thought out in detail” (p. 11).

In addition to focus groups, survey was used to collect qualitative data to address the three research questions. Because focus groups provide data on only a limited number of participants, survey added breadth to the qualitative data by providing respondents with a comment box to further explain their quantitative responses on 30 survey items on grading practices and influences of grading. As shown in Figure 2, focus groups and qualitative data from the survey were used to address Research Question 3 and also to triangulate quantitative results on Research Questions 1 and 2. Hesse-Biber (2010) described the importance of adding qualitative data to collected quantitative data for triangulation to enrich conclusions and increase credibility. It is the researcher’s intent that the pairing of focus group data with quantitative and qualitative survey data will add to the depth of the study, thereby increasing the validity of results.

Participants/Sample

This study was set in California’s South San Joaquin Valley, which is located in southern Central California. For the purpose of this study, the South San Joaquin Valley is defined as Fresno, Kern, and Tulare counties (Beacon Economics, 2014). The region has a population of 2.27 million residents and a

16% growth rate over the last decade (Beacon Economics, 2014). The region's diverse population is 54.4% Hispanic, 33.4% White, 5.8% Asian, and 3.8% African American. The region also has a 25.0% poverty rate, which is higher than the 15.9% rate of California (United States Census Bureau). Economically, the region is known for its vast agriculture, as it is responsible for 33.2% of California's total economic agricultural output in 2013, and also its prominent oil production (Beacon Economics, 2014).

The population considered in this study is high school students of the South San Joaquin Valley, and it is represented by the sample district given the pseudonym Sunnyside School District (SSD). The sample district is composed of 18 comprehensive high schools, 5 alternative education campuses, 4 special education centers, 3 career technical education sites, 1 adult education center, and 1 charter school. SSD serves 37,318 students and employs 1,631 teachers (California Department of Education). The diversity and poverty rate of SSD's student body is similar to that of both the student population of South San Joaquin Valley and the state of California (Table 4). This study focused on the comprehensive high schools in the district of study, and all references hereafter pertain only to those comprehensive sites.

Only regular education teachers from each school who taught at least one class of English, foreign language, mathematics, science, or social studies were considered for the study. Special education teachers were excluded because their grading practices would likely differ significantly from the practices of regular education teachers. These five subject areas were selected as the focus of this grading study because the grading practices of these teachers may reasonably be expected to differ from the practices of teachers of nonacademic subjects, and the grades assigned to students in academic subjects have different consequences than

Table 4

A Comparison of High School Population Size, Diversity, and Poverty Rate by Level of Organization

| Level of Organization | Student Population Size | Student Population Diversity | Student Poverty Rate |
|--------------------------|-------------------------|--|----------------------|
| California | 1.95 million | 53.6% Hispanic 24.6% White 8.8% Asian 6.0% African American | 57.3% |
| South San Joaquin Valley | 141,503 | 67.1% Hispanic 20.6% White 4.6% Asian 4.2% African American | 69.0% |
| Sample District | 37,318 | 62.6% Hispanic 23.9% White 2.6% Asian 6.0% African American | 63.5% |

Note. Student poverty rate refers to the percentage of students who qualify for the National School Lunch Program.

grades assigned to students in nonacademic subjects (Cross & Frary, 1999; Frary et al., 1992, 1993).

Sampling Rationale

Convenience sampling was used to select the district of study, as the researcher is a teacher within the district. Best and Kahn (2006) explained that this type of nonprobability is commonly used by educational researchers because of inherent limitations of creating comparison groups in schools. Although convenience sampling often precludes high levels of generalizability (Best & Kahn, 2006), the similarity of the district of study to high schools throughout California and particularly those in the South San Joaquin Valley (Table 4)

provides a certain amount generalizability that may not be common in studies using convenience sampling. Beyond convenience, this district was selected for the study because it presented a unique opportunity in its large size and student diversity to investigate differences in poverty levels without confounding variables experienced when studying multiple districts. Additionally, because the researcher is a teacher within the district, an opportunity existed for better survey response rates and greater depth in focus group responses.

Comprehensive sampling was utilized to maximize the sample size to best understand the teacher grading decision-making process. This type of sampling was feasible because of the manageable size of the teacher population in the district and because of the resourcefulness that accompanies being an educator within the district. The study utilized only comprehensive school sites because these are traditional schools, and the grading practices of traditional schools likely differ from nontraditional schools. The sample of schools used in focus groups—planned as two schools within each school poverty level group—were selected by convenience sampling, as the biggest factors in selecting school sites were the willingness of the administrators and teachers within each site to participate.

Data Collection and Instrumentation

Data were collected to address the study's research questions through the use of a Web-based survey questionnaire and focus groups. Prior to the use of either instrument, this study was reviewed and approved by California State University, Fresno's (CSU, Fresno) Committee on the Protection of Human Subjects Institutional Review Board and SSD's Institutional Review Board. Upon approval of both review boards, approval for teacher participation first in the survey and later in focus groups was obtained from school site principals. Data

were then collected with the survey and after in focus groups. The details of the data collection and instrumentation are explained below.

Survey

The survey was a self-administered Web-based questionnaire (Appendix A) created by the researcher to investigate the effect of school poverty level on teacher grading practices and the influences of teacher grading (Domains 1 and 3 of the teacher grading decision-making framework) (Table 5). To best achieve that purpose, the instrument was constructed through the use of elements of survey instruments used in the grading studies of Cross and Frary (1999), Frary et al. (1992, 1993) McMillan (2001), McMillan et al. (2002), and McMillan and Lawson (2001). Specifics of the adaptations are described in each of the three survey parts in the sections below.

The survey was administered electronically through SurveyMonkey, and it was primarily composed of 36 close-ended items organized into three parts: *teacher background, grading practices, and factors that influence grading rationale*. The survey began with an introduction that introduced the researcher, explained the study, addressed potential privacy concerns, provided contact information, and explained and asked for informed consent (which served as the first item of the survey). Each school site was sent a separate, but identical Web-based survey for a total of 18 surveys utilized for data collection. Although each survey was identical in form, utilizing a different URL for each survey added organization to data collection and analysis.

Steps toward data collection began at the end of the 2014-15 school year, 3 months prior to actual data collection, when emails were sent to principals of each school site to introduce the study and ask for permission of each school site to participate in the Web-based survey questionnaire at the beginning of the

Table 5

Alignment of Research Questions, Teacher Grading Decision-Making Framework, Instruments, and Analysis

| Research Question | Teacher Grading Decision-Making Framework | Instrument | Analysis |
|---|---|--|--|
| 1) How does school poverty level affect teacher grading practices? | Domain 3: Grading Practices | Survey Part 2, Focus Group Questions 1-5, Survey Open-Ended Items 7-36 | ANOVA, Kruskal-Wallis, descriptive statistics, NVivo: constant comparative analysis for themes |
| 2) How does school poverty level affect the influences of teacher grading? | Domain 1: Influences of Grading | Survey Part 3, Focus Group Questions 1-5, Survey Open-Ended Items 7-36 | MANOVA, descriptive statistics, NVivo: constant comparative analysis for themes |
| 3) What rationale do teachers of different school poverty levels use to make grading decisions? | Domain 2: Grading Rationale | Focus Group Questions 2-5, Survey Open-Ended Items 7-36 | NVivo: constant comparative analysis for emergent themes |

following school year. Three weeks into the 2015-16 school year, an email was again sent to each principal that explained the study and asked each principal to forward the survey link to their teachers targeted in the study. The emails also included an optional message for teachers that explained the purpose of the study. Between one and four reminders were given to principals, administrators, and teachers about the survey, depending on the response rate of each school site. Reminder emails were sent by the researcher, the researcher's administrators, and also a department chair at the researcher's school site. Reminder phone calls were

made by the researcher's school administrators on a small number of occasions, and face-to-face reminders were made by the researcher and the researcher's administrators. Although survey data were originally planned to be collected for 1 month, it was collected for a total of 2 months to obtain a larger sample size.

Part 1. Survey Part 1 contained three items on teacher background (Items 2 through 4) to establish a basis for comparison in data analysis. This section was developed from the first group of survey questionnaire items utilized by Cross and Frary (1999) and Frary et al. (1992, 1993) in their studies on teacher grading practices. Although the predominant basis for comparison was school poverty level (this was determined before data collection), additional data were collected in Part 1 to allow for comparison by years of teaching experience (less than 3, 3 to 6, 7 to 10, or more than 10), subject area (English, foreign language, mathematics, science, or social studies), and class level (half or more of classes college preparatory or fewer than half).

Part 2. Survey Part 2 addressed grading practices, which represents Domain 3 of the teacher grading decision-making process (Table 5). This section of the survey contained 18 items (Items 5 through 23). Items 5 and 6 asked teachers the percentage of A's and F's they assigned, respectively (less than 3%, 5 to 10%, 11 to 20%, or more than 20%). These two items were also adapted from Cross and Frary (1999) and Frary et al. (1992, 1993), and they were used to determine what difference, if any, existed in the distribution of assigned grades between schools of different poverty levels. Randall and Engelhard (2010) explained that while schools of varying poverty levels show differences in academic achievement measured by standardized tests, they rarely show these differences in report card grade distributions.

Items 7 through 23 were modified from factors considered by the survey questionnaires of McMillan (2001), McMillan and Lawson (2001), and McMillan et al. (2002) in their studies of teachers' assessment and grading practices. The questionnaires utilized by McMillan (2001), McMillan and Lawson (2001), and McMillan et al. (2002) each included a total of 34 items, 19 of which focused on factors used in grading practices and were modified to create Survey Part 2 in the present study.

The intent of the 17 items in Survey Part 2 was to determine how teachers conducted their grading practices so that they could be compared by school poverty level. These items were measured with a 5-point Likert-type scale ranging from *not at all* to *completely* to answer the extent to which second semester grades were based upon each factor or practice. This scale is similar to the 6-point scale used by McMillan et al. (2002), but it was modified to reflect the 5-point recommendation of Vagias (2006). Respondents were also provided with a comment box after Items 7 through 23 to include relevant comments if they chose to do so.

Part 3. Survey Part 3 addressed the influences of grading, which represents Domain 2 of the teacher grading decision-making process (Table 5). The researcher newly constructed this part of the survey questionnaire, as there was no existing instrument to measure influences of teacher grading as far as the researcher was aware. To be specific, the concept of influences of teacher grading (referred to as teacher knowledge, beliefs, expectations, and values; external factors; and classroom realities) was first developed by McMillan and Nash (2000) and refined by McMillan (2003) in their articles on teacher grading and assessment decision making. The data that McMillan and Nash (2000) and McMillan (2003) used to develop the concept and eventually the teacher grading

decision-making framework utilized in this study were collected through teacher interviews. To measure the influences of grading described by McMillan (2003) and McMillan and Nash (2000), the researcher developed 13 questionnaire items (Items 24 through 36). The explicit intent of these items was to assess the influence of teacher knowledge, beliefs, expectations, and values; external factors; and classroom realities on teacher decision-making rationale, and ultimately, on grading practices. When answering this final section of the survey, teachers selected from a 5-point Likert-type scale ranging from *not at all* to *completely* to answer the extent to which final first quarter grades were influenced by each factor. This scale also came from the recommendations of Vagias (2006). Like Part 2 of the questionnaire, respondents were also provided with a comment box after each item to include relevant comments if they chose to do so.

Focus Groups

Focus group interviews were utilized for data collection (Appendix B) in the second phase of the transformative explanatory sequential mixed methods design (Creswell, 2014), and they were the main instrument for qualitative data collection. The focus groups were utilized to address teacher grading rationale (Research Question 3 and Domain 2) and also to triangulate data on teacher grading practices (Research Question 1 and Domain 3) and influences of teacher grading (Research Question 2 and Domain 1) (Table 5). Focus group questions were developed from the focus group method recommendations of Morgan (1997) and were modified based on the survey results. Data were collected in an attempt to establish a thick description of the phenomenon of study and with the intent of reaching data saturation.

A total of four focus groups were conducted: one in a low-poverty school, one in a mid-poverty school, and two in high-poverty schools. Each focus group

consisted of two to five regular education teachers of English, foreign language, mathematics, science, and social studies in each group, and they lasted from 40 to 60 minutes each. The researcher conducted each focus group by utilizing an interview protocol (Appendix B). Each focus group was recorded with an electronic recording device, and the researcher recorded field notes according to an observational protocol (Appendix C).

Each session began with an introduction that introduced the research, explained the purpose of the study and the focus group, addressed possible concerns regarding confidentiality, and explained and obtained informed consent. Focus groups were semi-structured in nature and consisted of five key questions; however the researcher occasionally varied from the protocol to ask for a topic to be further explored or to redirect participants to the key questions (Table 6). The researcher transcribed each focus group himself within 48 hours of its completion.

Pilot Study

Both instruments were piloted prior to data collection, and results of the pilot study were used to refine each instrument. The pilot survey was conducted in a high school in a neighboring school district. After obtaining approval from the principal, the Web-based questionnaire was sent to the staff inviting all regular education teachers of at least one section of English, foreign language, mathematics, science, or social studies to complete the survey, which contained 37 closed-response items and two open-response items. The first open-response item asked respondents to describe the level of clarity of the 37 closed-response items, and the second open-response item asked respondents to describe the survey's comprehensiveness in listing factors utilized by teachers in grading practices and factors that influence grading practices. The pilot survey results prompted a number of revisions to the survey instrument, including the rewording of three

Table 6

Alignment of Focus Group Questions by Research Question and Teacher Grading Decision-Making Framework Domain

| Focus Group Question | Research Question | Teacher Grading Decision-Making Framework Domains |
|---|-------------------|---|
| 1) What are some challenges you face in creating student report card grades? | 1, 2, & 3 | 1, 2, & 3 |
| 2) How do you create student report card grades? | 1, 3 | 2, 3 |
| 3) How do your own knowledge, beliefs, and/or values (personal or professional) influence the decisions you make in grading? | 2, 3 | 1, 2 |
| 4) How do external factors (i.e., state testing, district policy, parents, administrators) influence the decisions you make in grading? | 2, 3 | 1, 2 |
| 5) How do classroom realities (i.e., social promotion, student absenteeism, disruptive behavior, and differing student ability within classes) influence the decisions you make in grading? | 2, 3 | 1, 2 |

items to enhance clarity, the deletion of an item on social promotion because a perceived irrelevance in the high school context, and the inclusion of optional comment boxes in a majority of items to provide an opportunity for qualitative data. The resulting finalized survey is presented in Appendix A.

A pilot focus group was conducted within the district of study to assess the effectiveness of the five focus groups questions and the interview and observational protocols in addressing the three research questions. It consisted of five total regular education teachers of English, foreign language, mathematics, science, and social studies, and all participants had previously taken the grading

survey. The focus group lasted for approximately 35 minutes. As a result of participant responses and feedback, all focus group questions were modified to clearly define the meaning of the term *grading*.

Basis of Comparison

School poverty level provided the primary basis of comparison in this study. Three poverty groups were created: *low-poverty schools*, *mid-poverty schools*, and *high-poverty schools*. *Low-poverty schools* were defined as schools with 50% or less of students eligible for FRPL, while *mid-poverty schools* were defined as schools with 50.1 to 75% of students eligible for FRPL, and *high-poverty schools* were defined as schools with more than 75% of students eligible for FRPL. The three poverty groups established for this study were based upon the four school poverty levels established by the U.S. Department of Education: *low-poverty schools* are defined as public schools in which 25.0% or less of students are eligible for FRPL, *mid-low poverty schools* are defined as public schools in which 25.1 to 50.0% of students are eligible for FRPL, *mid-high poverty schools* are defined as public schools in which 50.1 to 75.0% of students are eligible for FRPL, and *high-poverty schools* are defined as public schools in which more than 75.0% of students are eligible for FRPL (Kena et al., 2015). Three groups were created instead of four because of concerns of greatly disproportionate comparison groups. To address this concern, groups designated as *low poverty* and *mid-low poverty* by the Department of Education were merged into a single *low-poverty* group, while the group designated as *mid-high poverty* by the Department of Education was termed *mid poverty* in this study. The *high-poverty* group remained unchanged. Although these changes addressed some of the concerns for disproportionality in samples across groups, the high-poverty group remained much larger than the two other groups. However, this larger

group is representative of the disproportionate number of high-poverty schools in both the district of study and the South San Joaquin Valley.

Table 7 shows the schools and populations that make up each school poverty level group. Eight schools in the study were classified as high poverty, five schools were classified as mid poverty, and five schools were classified as low poverty. As shown in Table 7, the schools in the high-poverty group greatly outnumber the schools in the two other groups in total schools and total population. These high numbers do reflect the high number of underprivileged students in SSD.

Table 7

School Poverty Level Groups

| School Poverty Level | | | | | |
|----------------------|-----|---------|-----|---------|-----|
| Low | | Mid | | High | |
| Schools | N | Schools | N | Schools | N |
| I | 50 | H | 71 | G | 60 |
| L | 57 | N | 50 | J | 60 |
| O | 49 | P | 13 | K | 51 |
| Q | 50 | S | 37 | M | 59 |
| W | 53 | T | 58 | R | 60 |
| -- | -- | -- | -- | U | 36 |
| -- | -- | -- | -- | V | 51 |
| -- | -- | -- | -- | X | 50 |
| Total | 259 | -- | 229 | -- | 427 |

Note. N = the number of regular education teachers of English, foreign language, mathematics, science, or social studies in each school.

Data Analysis

Quantitative Data

In this causal-comparative component of the study, data were analyzed with descriptive and inferential statistical analyses with the use of SPSS 23.0.

Descriptive statistics in the form of means, medians, and standard deviations were calculated to explore results to “determine general trends in the data” (Creswell & Plano Clark, 2011, p. 206). Additionally, descriptive statistics were used to describe teacher grading practices and grading decision making by comparison group. Both parametric and non-parametric inferential statistics were used to analyze quantitative survey data on grading practices and influences of grading. One-way analyses of variance (ANOVAs) were used to compare means of grade distributions and subjective grading practices by school poverty level, while a multivariate analysis of variance (MANOVA) was used to compare 13 influences of teacher grading by poverty level. A MANOVA was selected to analyze the 13 influences because the dependent variables displayed a conceptual relationship best tested by the MANOVA (Leech, Barrett, & Morgan, 2011). Kruskal-Wallis tests were used to compare each of the 17 grading practices in the survey by school poverty level because unequal variances across groups and non-normal dependent variable distributions eliminated the possibility of the one-way ANOVA (Morgan, Leech, Gloeckner, & Barrett, 2011).

Data analysis commenced immediately upon the close of the survey by calculating survey frequencies and response rates and also by determining means, medians, and standard deviations for each of the dependent variables. After calculating descriptive statistics, one-way ANOVAs were used to compare teacher reported percentages of assigned A's and F's by school poverty level. Next, Kruskal-Wallis tests were used to compare the use of the 17 teacher grading

practices by school poverty level. To further analyze teacher grading practices, a subjective grading index (SGI) score, similar in concept to the one used by Goldwater and Nutt (1999), was calculated for each of the survey participants. These scores represented the subjectivity of each teacher's grading practices based on their responses to the 17 grading practice items in the survey and the recommendations of grading researchers (Cross & Frary, 1999; McMillan, 2001; McMillan & Lawson, 2001; McMillan et al., 2002). Scores were calculated by averaging the numeric responses from Items 7 through 23 from Survey Part 2. However, although 15 of the items were considered subjective practices, two practices (Items 8 and 16) were considered objective (Cross & Frary, 1999; McMillan, 2001; McMillan & Lawson, 2001; McMillan et al., 2002), and they were reverse coded for data analysis purposes. A final SGI score of 1.0 represented the minimum grading subjectivity in the index (and maximum objectivity), while a score of 5.0 represented the maximum subjectivity. One-way ANOVAs compared SGI scores by school poverty level. Finally, a MANOVA was used to analyze the effect of school poverty level on the 13 assessed influences of teacher grading.

Qualitative Data

Qualitative data were analyzed with a constant comparative analysis method (Corbin & Strauss, 2008). Focus group data were transcribed by the researcher and combined with the survey open-response data in NVivo 11.0. The raw qualitative data underwent three rounds of coding: open, axial, and selective coding. Analysis began with open coding to gain a general understanding of the data and to create data categories. Axial coding followed to reconfigure and better organize emergent categories (Saldana, 2012). During the process of axial coding, one category was designated as the *core phenomenon*, as it represented a central

concept found in the data (Corbin & Strauss, 2008). The reorganization of categories during axial coding was conducted around the core phenomenon. The final round of coding, selective coding, was done by interpreting emergent themes from the refined categories along with McMillan's (2003) teacher grading decision-making model (Corbin & Strauss, 2008; Creswell, 2007).

Data Mixing

Quantitative and qualitative data were mixed to address Research Questions 1 and 2 and the guiding research question. In addressing Research Question 1, on the effect of school poverty level on teacher grading practices, the resulting themes from qualitative data analysis were used to triangulate quantitative results from data collected with Survey Parts 1 and 2. In addressing Research Question 2, on the effect of school poverty level on the influences of teacher grading, the resulting themes from qualitative data analysis were used to triangulate quantitative results from data collected with Survey Parts 1 and 3. Finally, in addressing the guiding research question, on the effects of school poverty level on the teacher grading decision-making process, the resulting themes produced from qualitative data analysis were used to triangulate quantitative results from data collected with Survey Parts 1 through 3.

Limitations

This study includes a number of limitations. The most significant of these apply to external validity. Because the study used a convenience sample, results may be limited in their generalizability (Best & Kahn, 2006). However as previously described, the sample district is representative of the South San Joaquin Valley. To address the additional threat to external validity, generalizations were not made to schools outside of this region. Also, the sample of participants in

focus groups—for both the school sites and teachers within the sites—was determined upon accepted invitation. These four sample schools and the two to five teachers within each school that participated were not random, and participants may “have higher morale, less insecurity, greater willingness to try a new approach, and a greater desire to improve their performance” (Best & Kahn, 2006, p. 176). This limitation was only addressed by stating its existence.

The study also faced a number of threats to internal validity. Experimenter effects may have played a role in data collection, as the researcher is a teacher in the sample district (Ary, Jacobs, & Sorenen, 2006). To address this limitation, interview and observational protocols were used in focus groups. Additionally, an inherent limitation in focus group interviews is social desirability bias (Ary et al., 2006; Bradburn et al., 2004). To combat this issue, the researcher utilized the recommendations of Bradburn et al. (2004) of establishing rapport with the respondents, doing his best to put all members at their ease, and appearing nonjudgmental throughout the process. It should be noted that this particular threat may also apply to the survey questionnaire (Ary et al., 2006). Finally, because the survey questionnaire was self-reported, the accuracy of responses could not be verified. This limitation was only addressed by stating its existence.

CHAPTER 4: RESULTS

Introduction

This study investigated the effects of school poverty level on the teacher grading decision-making process. In doing so, an overarching critical pedagogy theoretical framework was used to provide a focusing lens on the impact of school poverty level on grading. Additionally, a second theoretical framework—McMillan's (2003) teacher grading decision-making model—was used to construct essential structural components of the study in the form of research questions and data collection instruments. The study utilized a single guiding research question: How does school poverty level affect the teacher grading decision-making process?

This guiding question was broken into three specific questions to address each of the domains of the teacher grading decision-making framework:

1. How does school poverty level affect teacher grading practices?
2. How does school poverty level affect the influences of teacher grading?
3. What rationale do teachers of different school poverty levels use to make grading decisions?

Utilizing a transformative explanatory sequential mixed methods design (Creswell, 2014), a 36-item Web-based survey questionnaire was used to collect quantitative and qualitative data to explain the impact of school poverty level on teacher grading decision making. The survey was principally designed to collect quantitative data, and all 36 items were used for some form of quantitative data analysis. Additionally, Items 7 through 36 each included an optional open-ended response for teachers to provide comments to the close-ended response items, and

the qualitative data were later combined with focus group data for analysis. The survey was piloted at a high school in a neighboring school district, and results were used to make multiple revisions to the instrument. The final survey was sent to regular education teachers of English, foreign language, mathematics, science, and social studies at all 18 comprehensive high schools within the district of study.

Descriptive statistics from the quantitative data were used to refine the second instrument, which was the interview protocol used to conduct teacher focus group interviews. This instrument was tested in a pilot focus group at a site within the district, and results were used to refine the instrument. The refined instrument was utilized in four focus groups throughout the district: one in a low-poverty school, one in a mid-poverty school, and two in high-poverty schools. Each focus group was composed of two to five regular education teachers of core subjects, and teachers were asked five questions in a semi-structured format. All qualitative data, including that collected through the survey and focus groups, were used to triangulate quantitative data on Research Questions 1 and 2 and to fully address Research Question 3.

Quantitative data were analyzed with descriptive analyses and both parametric and nonparametric tests. One-way analysis of variance (ANOVA) tests were used to compare the means of assigned A's, assigned F's, and subjective grading index (SGI) scores (created from the 17 grading practices survey data) by school poverty level. Kruskal-Wallis tests were used to compare the means for each of the 17 different teacher grading practices by school poverty level, as unequal variances and non-normal distributions prevented the use of ANOVA tests (Morgan et al., 2011). A multiple analysis of variance (MANOVA) test was used to compare the means of the 13 influences of grading by school poverty level because unlike the 17 teacher grading practices, the 13 influences displayed a

conceptual relationship between the dependent variables best tested by the MANOVA (Leech et al., 2011).

Qualitative data, in the form of survey open-response Items 7 through 36 and focus group data, were analyzed with a constant comparative analysis method (Corbin & Strauss, 2008). All qualitative data were combined and entered into NVivo version 11.0 for three separate rounds of coding. Analysis began with open coding to create categories, which was followed by axial coding to consolidate categories. In this stage of coding, one category was determined to be a *core phenomenon*, and all other categories were constructed around this phenomenon (Corbin & Strauss, 2008). Finally, in selective coding, themes were created from the categories that aligned with McMillan's (2003) teacher grading decision-making model (Corbin & Strauss, 2008; Creswell, 2007).

Demographics of the Sample

The key demographics of the 18 high schools within SSD display high diversity throughout the district. Table 8 shows a comparison of the demographic information by school site. Although a majority of the schools within the district (13 of 18) have a student body in which the largest ethnic group is Hispanic, school ethnic diversity indices, which measure the distribution of student race/ethnicity categories (0 being the lowest and 100 being the highest), range from 4 to 57. Differences also include a wide school minority rate range of 39% to 96.9%. While 8.8% of all students are classified as English Learners, sites range on this figure from 0.8% to 20.8%. Schools throughout the district also vary by enrollment and school poverty level. School enrollment ranges from 469 to 2,790 students, and school poverty rates within the district—the percentage of students within a school that qualify for FRPL—range from 17.4% to 89.5%. On

average, 63.5% of students qualify for FRPL, although the district shows a vast range of 17.4% to 89.5% (California Department of Education).

Of the 1,631 teachers in SSD, 74% are White, 17% are Hispanic, 3% are African American, and 3% are Asian. On average, teachers have a total of 15 years of service, while the average experience within the district is 13 years. Approximately 9% of SSD teachers are in their first 2 years of teaching (California Department of Education).

Quantitative Data Analysis

Quantitative data were the predominant data type used to address Research Questions 1 and 2 to determine the effects of school poverty level on teacher grading practices and on the influences of teacher grading. Qualitative data were later used to triangulate these quantitative results. Data were collected with a survey, and all regular education teachers of at least one period of English, foreign language, mathematics, science, or social studies were invited to complete the survey. The total study population consisted of 915 teachers from schools across the district, while the population of each site ranged from 13 to 71. Table 9 displays survey response data for each school site. It should be noted that the core subject teachers targeted in this study represented 56% of the total population of teachers in the district. Of the 915 teachers in the targeted study population, 325 teachers (36%) began the survey, while 251 completed the survey for a response rate of 27.4%.

Survey responses varied greatly by school poverty level group, as displayed in Table 10. Notably, despite the disproportionately high number of schools (8) and teachers (464) in the high-poverty group, the sample sizes of the three groups are similar. The response rate was highest in mid-poverty schools (44.7%) and lowest in high-poverty schools (18.1%). A major reason for the low response rate in the

Table 8

Demographic Information by School

| School | Enrollment | FRPL (%) | Largest Ethnic Group | EDI | EL (%) |
|--------|------------|----------|----------------------|-----|--------|
| G | 2,473 | 89.2 | Hispanic | 13 | 20.8 |
| H | 2,790 | 64.2 | Hispanic | 25 | 8.2 |
| I | 1,969 | 22.9 | White | 4 | 1.5 |
| J | 2,114 | 82.3 | Hispanic | 15 | 12.2 |
| K | 1,857 | 82.3 | Hispanic | 21 | 11.7 |
| L | 2,324 | 23.6 | White | 18 | 0.8 |
| M | 2,418 | 79.0 | Hispanic | 38 | 10.0 |
| N | 1,966 | 62.7 | Hispanic | 11 | 8.1 |
| O | 2,048 | 49.6 | Hispanic | 32 | 2.9 |
| P | 469 | 59.9 | White | 37 | 0.9 |
| Q | 1,926 | 17.4 | White | 42 | 1.3 |
| R | 2,004 | 89.2 | Hispanic | 35 | 15.7 |
| S | 1,585 | 71.6 | White | 24 | 2.1 |
| T | 2,278 | 68.9 | Hispanic | 49 | 5.9 |
| U | 1,440 | 76.0 | Hispanic | 57 | 20.7 |
| V | 1,967 | 89.5 | Hispanic | 39 | 14.3 |
| W | 2,087 | 31.1 | Hispanic | 41 | 1.6 |
| X | 2,124 | 78.7 | Hispanic | 36 | 10.2 |

Note. FRPL = Free or Reduced-Price Lunch; EDI = ethnic diversity index; EL = English Learners.
Adapted from California Department of Education.

Table 9

Survey Responses by School

| School | N | Surveys Started | Surveys Completed | Completion Rate (%) | Response Rate (%) |
|--------|-----|-----------------|-------------------|---------------------|-------------------|
| G | 60 | 16 | 13 | 81.3 | 21.7 |
| H | 71 | 27 | 20 | 74.1 | 28.2 |
| I | 50 | 20 | 18 | 90.0 | 36.0 |
| J | 60 | 11 | 9 | 81.8 | 15.0 |
| K | 51 | 22 | 16 | 72.7 | 31.4 |
| L | 57 | 23 | 17 | 73.9 | 29.8 |
| M | 59 | 14 | 10 | 71.4 | 16.9 |
| N | 50 | 41 | 37 | 90.2 | 74.0 |
| O | 49 | 17 | 15 | 88.2 | 30.6 |
| P | 13 | 10 | 8 | 80.0 | 61.5 |
| Q | 50 | 21 | 14 | 66.7 | 28.0 |
| R | 60 | 6 | 4 | 66.7 | 6.7 |
| S | 37 | 11 | 10 | 90.9 | 27.0 |
| T | 58 | 16 | 11 | 68.8 | 19.0 |
| U | 36 | 0 | 0 | 0.0 | 0.0 |
| V | 51 | 30 | 21 | 70.0 | 41.2 |
| W | 53 | 23 | 17 | 73.9 | 32.1 |
| X | 50 | 17 | 11 | 65.0 | 22.0 |
| Total | 915 | 325 | 251 | 77.2 | 27.4 |

Note. N = the number of regular education teachers of English, foreign language, mathematics, science, or social studies in each school.

high-poverty schools group was because of the exceptionally low rate of Site R (6.7%) and the non-participation of Site U. All subsequent data will only reflect the data from the 17 participating sites.

Table 10

Survey Responses by School Poverty Level

| | School Poverty Level | | |
|-------------------|----------------------|------|------|
| | Low | Mid | High |
| Target Population | 259 | 192 | 464 |
| Completed Surveys | 81 | 86 | 84 |
| Response Rate (%) | 31.3 | 44.7 | 18.1 |

Unlike the great range of survey completion by school site and group, the types of teachers that responded to the survey were remarkably similar. Table 11 shows the teaching experience, major teaching assignment, and primary class level of teachers who responded to the survey by school poverty level. Across poverty levels, the majority of teachers taught for more than 10 years. English teachers made up the largest proportion of respondents in low- (37%) and high-poverty schools (42.9%), while mathematics teachers (33.7%) represented a slightly higher proportion of respondents than English teachers (30.2%) in mid-poverty schools. Foreign language teachers represented the smallest proportion of teachers in all groups, and the vast majority of classes taught by respondents were college preparatory. However, the proportion of teachers with fewer than 50% of classes of college preparatory level was higher in high-poverty schools (22.6%) as compared to mid-poverty schools (11.6%) and low-poverty schools (8.6%).

Table 11

Demographics of Survey Participants by School Poverty Level

| Variable | School Poverty Level | | | | | |
|----------------------------------|----------------------|------|-----------|------|-----------|------|
| | Low | | Mid | | High | |
| | Frequency | % | Frequency | % | Frequency | % |
| Teaching Experience | | | | | | |
| < 3 | 1 | 1.2 | 2 | 2.3 | 3 | 3.6 |
| 3 to 6 | 5 | 6.2 | 13 | 15.1 | 11 | 13.1 |
| 7 to 10 | 6 | 7.4 | 6 | 7.0 | 11 | 13.1 |
| > 10 | 69 | 85.2 | 65 | 75.6 | 59 | 70.2 |
| Major Teaching Assignment | | | | | | |
| English | 30 | 37.0 | 26 | 30.2 | 36 | 42.9 |
| Foreign Language | 4 | 4.9 | 3 | 3.5 | 3 | 3.6 |
| Mathematics | 16 | 19.8 | 29 | 33.7 | 13 | 15.5 |
| Science | 16 | 19.8 | 23 | 26.7 | 16 | 19.0 |
| Social Studies | 15 | 18.5 | 14 | 16.3 | 16 | 19.0 |
| Class Level | | | | | | |
| ≥ 50% CP | 74 | 91.4 | 76 | 88.4 | 65 | 77.4 |
| < 50% CP | 7 | 8.6 | 10 | 11.6 | 19 | 22.6 |

Note. CP = College preparatory classes.

Grading Practices

The first aspect of teacher grading practices that was investigated was the proportion of A's and F's assigned by teachers by school poverty level. Table 12 illustrates the mean, standard deviation, and median of teacher self-reported assigned A's and F's at each school poverty level. Median scores were included in Table 12, as well in all subsequent descriptive statistic tables of grading practices because of the high variance of teacher self-reported grading practices within comparison groups (Morgan et al., 2011). Teacher scores were self-reported based on a four-item scale in which 1 = less than 3%, 2 = 5 to 10%, 3 = 11 to 20%, and 4 = more than 20%.

Table 12

Means, Standard Deviations, and Medians Comparing Teacher Assigned Grades by School Poverty Level

| Grading Practice | School Poverty Level | | | | | | | | |
|----------------------|----------------------|-----------|------------|----------|-----------|------------|----------|-----------|------------|
| | Low | | | Mid | | | High | | |
| | <i>M</i> | <i>SD</i> | <i>Mdn</i> | <i>M</i> | <i>SD</i> | <i>Mdn</i> | <i>M</i> | <i>SD</i> | <i>Mdn</i> |
| Teacher Assigned A's | 3.05 | .88 | 3.00 | 2.58 | .91 | 3.00 | 2.63 | .92 | 3.00 |
| Teacher Assigned F's | 1.59 | .70 | 1.00 | 2.01 | .91 | 2.00 | 1.83 | .92 | 2.00 |

As shown in Table 12, teachers of low-poverty schools reported assigning the highest percentage of A's ($M = 3.05$), near the *11 to 20%* category, and the lowest percentage of F's ($M = 1.59$)—somewhere between the *less than 3%* and *5 to 10%* categories. Mean scores (2.58) of self-reported assigned A's in mid-poverty schools were similar to the mean scores (2.63) in high-poverty schools, which falls somewhere between the *5 to 10%* and *11 to 20%* categories. Teachers of mid-poverty schools also displayed the highest mean (2.01) for assigned F's, at approximately *5 to 10%*.

A one-way ANOVA was conducted to test for significant differences between school poverty levels in mean scores of teacher assigned A's and F's. A statistically significant difference was found between the three school poverty levels on self-reported teacher assigned A's, $F(2, 248) = 6.68, p = .001$ and on self-reported teacher assigned F's, $F(2, 248) = 5.06, p = .007$. Complete results are shown in Table 13.

To determine pairwise contrasts in the ANOVA results shown in Table 13, Post Hoc Tukey HSD tests were conducted (Morgan et al., 2011). Results of post hoc tests showed significant differences in teacher assigned A's between low- and mid-poverty schools with a medium effect size according to Cohen (1988) ($p < .01, d = .53$). Additionally, significant differences were found in teacher assigned

Table 13

ANOVA for Teacher Assigned Grades by School Poverty Level

| Factor | Source | SS | df | MS | F | p |
|--------------|----------------|--------|-----|------|------|--------|
| Assigned A's | Between Groups | 10.90 | 2 | 5.45 | 6.68 | .001** |
| | Within Groups | 202.29 | 248 | .82 | | |
| | Total | 213.19 | 250 | | | |
| Assigned F's | Between Groups | 7.36 | 2 | 3.68 | 5.06 | .007** |
| | Within Groups | 180.21 | 248 | .727 | | |
| | Total | 187.57 | 250 | | | |

Note. ** $p < .01$

A's between low- and high-poverty schools with a medium effect size ($p < .01$, $d = .47$). Significant differences were also found in teacher assigned F's between low- and mid-poverty schools with a medium effect size according to Cohen (1988 ($p < .01$, $d = .52$).

The second aspect of grading practices investigated was the extent to which teachers of each school poverty level utilized 17 different grading practices to create report card grades. Table 14 illustrates the means, standard deviations, and medians of teacher responses to the 17 grading practices assessed in the study. Teacher scores were self-reported based on a five-item Liker-type scale in which 1 = not at all, 2 = slightly, 3 = somewhat, 4 = largely, and 5 = completely. Across all school poverty levels (low, mid, and high), Table 14 displays the highest mean responses on *student academic achievement* (3.84, 4.06, 4.00), *specific learning objectives mastered by students* (3.78, 3.92, 3.85), and *student ability level* (3.53, 3.63, 3.58). Lowest means were seen on *grade distributions of other teachers* (1.20, 1.20, 1.08), *student performance compared to students from previous years* (1.22, 1.31, 1.31), and *student disruptive behavior/conduct* (1.23, 1.36, 1.40).

Table 14

Means, Standard Deviations, and Medians Comparing Teacher Grading Practices by School Poverty Group

| Grading Practices | School Poverty Level | | | | | | | | |
|--|----------------------|-----------|------------|----------|-----------|------------|----------|-----------|------------|
| | Low | | | Mid | | | High | | |
| | <i>M</i> | <i>SD</i> | <i>Mdn</i> | <i>M</i> | <i>SD</i> | <i>Mdn</i> | <i>M</i> | <i>SD</i> | <i>Mdn</i> |
| 1) Student Ability Level | 3.53 | .81 | 4.00 | 3.63 | .99 | 4.00 | 3.58 | .81 | 4.00 |
| 2) Student Academic Achievement | 3.84 | .78 | 4.00 | 4.06 | .76 | 4.00 | 4.00 | .62 | 4.00 |
| 3) Student Disruptive Behavior/Conduct | 1.23 | .55 | 1.00 | 1.36 | .68 | 1.00 | 1.40 | .71 | 1.00 |
| 4) Student Effort | 3.09 | .90 | 3.00 | 3.00 | 1.11 | 3.00 | 3.08 | 1.02 | 3.00 |
| 5) Student Participation and/or Paying Attention | 2.35 | 1.08 | 2.00 | 2.43 | 1.09 | 2.00 | 2.58 | 1.11 | 3.00 |
| 6) Student Improvement of Performance | 2.60 | .93 | 3.00 | 2.71 | .94 | 3.00 | 2.80 | .99 | 3.00 |
| 7) Grade Distributions of Other Teachers | 1.20 | .66 | 1.00 | 1.20 | .65 | 1.00 | 1.08 | .35 | 1.00 |
| 8) Student Performance of other Students in Classes | 1.49 | .82 | 1.00 | 1.45 | .79 | 1.00 | 1.59 | .91 | 1.00 |
| 9) Student Performance Compared to Students from Previous Years | 1.22 | .63 | 1.00 | 1.31 | .74 | 1.00 | 1.31 | .78 | 1.00 |
| 10) Specific Learning Objectives Mastered by Students | 3.78 | .88 | 4.00 | 3.92 | .75 | 4.00 | 3.85 | .72 | 4.00 |
| 11) Formal or Informal School or District Policy on Grade Distributions | 1.44 | 1.00 | 1.00 | 1.48 | 1.07 | 1.00 | 1.62 | 1.12 | 1.00 |
| 12) Student Effort, Improvement, Behavior and/or Other Non-Test Indicators for Borderline Grades | 2.25 | .92 | 2.00 | 2.47 | .95 | 2.00 | 2.58 | 1.02 | 3.00 |

Table 14 (cont.)

| Grading Practices | School Poverty Level | | | | | | | | |
|--|----------------------|-----------|------------|----------|-----------|------------|----------|-----------|------------|
| | Low | | | Mid | | | High | | |
| | <i>M</i> | <i>SD</i> | <i>Mdn</i> | <i>M</i> | <i>SD</i> | <i>Mdn</i> | <i>M</i> | <i>SD</i> | <i>Mdn</i> |
| 13) Student Completion of Homework | 1.93 | .79 | 2.00 | 1.92 | .90 | 2.00 | 2.01 | 1.01 | 2.00 |
| 14) Quality of Student Completed Homework | 2.83 | .96 | 3.00 | 2.33 | 1.05 | 2.00 | 2.76 | 1.10 | 3.00 |
| 15) Inclusion of Zeros for Incomplete Assignments or Assessments | 3.20 | 1.13 | 3.00 | 3.26 | 1.29 | 3.00 | 3.05 | 1.24 | 3.00 |
| 16) Student Extra Credit for Academic Performance | 1.74 | .54 | 2.00 | 1.52 | .66 | 1.00 | 1.57 | .65 | 1.50 |
| 17) Student Extra Credit for Non-Academic Performance | 1.22 | .59 | 1.00 | 1.13 | .40 | 1.00 | 1.10 | .30 | 1.00 |

Across all school poverty levels (low, mid, and high), highest standard deviations were seen on *inclusion of zeros for incomplete assignments or assessments* (1.13, 1.29, 1.24) and *student participation and/or paying attention* (1.08, 1.09, 1.11). Lowest standard deviations were seen on *student extra credit for academic performance* (.54, .66, .65) and *student extra credit for non-academic performance* (.59, .40, .30). Notably, a number of the standard deviations of group item scores differed greatly from the other comparison groups including *grade distribution of other teachers in high-poverty schools* (.35) and *student disruptive behavior/conduct in low-poverty schools* (.55), which were lower than their respective comparison groups, and *student extra credit for non-academic performance in low-poverty schools* (.59), which was higher than comparison groups (Table 14).

Low-poverty schools displayed notably high between-group means for *student extra credit for academic performance* (1.74) and *student extra credit for non-academic performance* (1.22), while the high-poverty school group displayed notably high between-group means for *student disruptive behavior/conduct* (1.40); *student participation and/or paying attention* (2.58); *student improvement of performance* (2.80); *student performance of other students in classes* (1.59); *formal or informal school or district policy on grade distributions* (1.62); and *student effort, improvement, behavior and/or other non-test indicators for borderline grades* (2.58). The low-poverty school group displayed notably low between-group means for *student academic achievement* (3.84); *student disruptive behavior/conduct* (1.23); *student participation and/or paying attention* (2.35); and *student effort, improvement, behavior and/or other non-test indicators for borderline grades* (2.25). The high-poverty school group displayed notably low between-group means for *grade distributions of other teachers* (1.08) and *inclusion of zeros for incomplete assignments or assessments* (3.05) (Table 14).

A Kruskal-Wallis nonparametric test was conducted to test for significant differences between school poverty levels in the 17 different grading practices. This test was used instead of parametric alternatives such as the ANOVA and MANOVA because unequal variances across groups violated assumptions of the tests (Morgan et al., 2011). The Kruskal-Wallis test indicated that the three school poverty groups differed significantly on *quality of student completed homework*, $X^2(2, N = 251) = 11.03, p = .004$, and *student extra credit for academic performance*, $X^2(2, N = 251) = 8.30, p = .016$. Table 15 displays the complete results for each grading practice.

Table 15

Kruskal-Wallis Test Results of Grading Practices by School Poverty Group

| Grading Practice | X^2 | <i>df</i> | <i>p</i> |
|--|-------|-----------|----------|
| 1) Student Ability Level | 2.07 | 2 | .354 |
| 2) Student Academic Achievement | 4.45 | 2 | .108 |
| 3) Student Disruptive Behavior/Conduct | 2.98 | 2 | .225 |
| 4) Student Effort | .52 | 2 | .770 |
| 5) Student Participation and/or Paying Attention | 1.78 | 2 | .410 |
| 6) Student Improvement of Performance | 1.61 | 2 | .447 |
| 7) Grade Distributions of Other Teachers | 1.66 | 2 | .437 |
| 8) Student Performance of other Students in Classes | 1.20 | 2 | .549 |
| 9) Student Performance Compared to Students from Previous Years | .53 | 2 | .767 |
| 10) Specific Learning Objectives Mastered by Students | 1.48 | 2 | .476 |
| 11) Formal or Informal School or District Policy on Grade Distributions | 1.42 | 2 | .491 |
| 12) Student Effort, Improvement, Behavior and/or Other Non-Test Indicators for Borderline Grades | 4.57 | 2 | .102 |
| 13) Student Completion of Homework | .29 | 2 | .864 |
| 14) Quality of Student Completed Homework | 11.03 | 2 | .004** |
| 15) Inclusion of Zeros for Incomplete Assignments or Assessments | 1.24 | 2 | .538 |
| 16) Student Extra Credit for Academic Performance | 8.30 | 2 | .016* |
| 17) Student Extra Credit for Non-Academic Performance | 2.09 | 2 | .352 |

Note. * $p < .05$; ** $p < .01$

Mann-Whitney post hoc tests compared the three school poverty levels on grading practice 14: *quality of student completed homework* and 16: *student extra credit for academic performance* with a Bonferonni corrected p value of .017 to determine statistical significance (Morgan et al., 2011). For grading practice 14: *quality of student completed homework*, the mean rank for low-poverty schools (137.39, $n = 81$) was significantly higher than in teachers in mid-poverty schools (105.72, $n = 86$), $z = -3.11$, $p = .002$, $r = -.24$. This effect size may be interpreted as small to medium according to Cohen (1988). For the same grading practice, the mean rank for high-poverty schools (135.79, $n = 84$) was significantly higher than for mid-poverty schools (105.72, $n = 86$), $z = -2.62$, $p = .009$, $r = -.20$. This effect size may be interpreted as small to medium according to Cohen (1988). Table 16 displays the complete results of the post-hoc test for Grading Practice 14.

Table 16

Mann-Whitney U Test for Teacher Grading Practice 14: Quality of Student Completed Homework

| Comparison Groups | U | Z | p | r |
|----------------------------|---------|-------|-------|------|
| Low Poverty – Mid Poverty | 2551.50 | -3.11 | .002* | -.24 |
| Low Poverty – High Poverty | 3393.00 | -.03 | .975 | -- |
| Mid Poverty – High Poverty | 2799.00 | -2.62 | .009* | -.20 |

Note. * $p < .017$

For Grading Practice 16: *Student extra credit for academic performance*, the mean rank for low-poverty schools (142.70, $n = 81$) was significantly higher than in teachers in mid-poverty schools (115.07, $n = 86$), $z = -2.72$, $p = .006$, $r = -.17$. This effect size may be interpreted as small to medium according to Cohen

(1988). Table 17 displays the complete results of the post-hoc test for Grading Practice 16.

Table 17

Mann-Whitney U Test for Teacher Grading Practice 16: Student Extra Credit for Academic Performance

| Comparison Groups | <i>U</i> | <i>Z</i> | <i>p</i> | <i>r</i> |
|----------------------------|----------|----------|----------|----------|
| Low Poverty – Mid Poverty | 2723.500 | -2.72 | .006* | -.17 |
| Low Poverty – High Poverty | 2809.00 | -2.19 | .028 | -- |
| Mid Poverty – High Poverty | 3431.500 | -.63 | .526 | -- |

Note. * $p < .017$

An SGI score was created for each survey respondent by averaging survey responses to the 17 grading practices. Factor 2: *student academic achievement* and Factor 10: *specific learning objectives mastered by students*—the only two researcher-recommended grading practices—were reverse coded to align with the scale of the other grading practices (Cross & Frary, 1999). This resulted in an index in which a score of 1.0 represents minimum grading objectivity and 5.0 represents maximum grading subjectivity. A Cronbach's alpha score was calculated to assess the internal consistency reliability of the 17 combined grading practices that produced the SGI score (Table 18), and the calculated score ($\alpha = .66$) was found to be below the minimum desired score ($\alpha = .70$). The α score ($\alpha = .66$) was based on standardized items because of the large variance in grading practice means and standard deviations (Morgan et al., 2011).

Table 18

Cronbach's Alpha for Teacher Grading Practices

| <i>N</i> | <i>Items</i> | <i>Reliability (α)^a</i> |
|----------|--------------|--|
| 251 | 17 | .66 |

Note. Based on standardized items.

The mean SGIs of all school poverty levels were similar, although low-poverty schools displayed a somewhat low standard deviation (.26) and mid-poverty schools showed a somewhat low median (2.03). Table 19 displays the complete results of the descriptive statistics of SGIs by school poverty level.

Table 19

Means, Standard Deviations, and Medians Comparing Subjective Grading Index by School Poverty Level

| <i>School Poverty Level</i> | <i>M</i> | <i>SD</i> | <i>Mdn</i> |
|-----------------------------|----------|-----------|------------|
| Low | 2.10 | .26 | 2.12 |
| Mid | 2.07 | .39 | 2.03 |
| High | 2.14 | .35 | 2.18 |

A one-way ANOVA was conducted to test for significant differences in SGI scores between school poverty levels. No statistical significant differences were found. The complete results are shown in Table 20.

Table 20

ANOVA Results for Subjective Grading Index by School Poverty Group

| | <i>SS</i> | <i>df</i> | <i>MS</i> | <i>F</i> | <i>p</i> |
|----------------|-----------|-----------|-----------|----------|----------|
| Between Groups | .19 | 2 | .093 | .806 | .448 |
| Within Groups | 28.77 | 248 | .116 | | |
| Total | 28.95 | 250 | | | |

Influences of Teacher Grading

Influences of teacher grading were measured and analyzed by school poverty level with the use of 13 survey items, which consisted of Items 24 to 36 in the survey. Table 21 illustrates the mean, standard deviation, and median for the 13 influences assessed in the study. Teacher scores in Table 21 were self-reported based on a five-item Likert-type scale in which 1 = not at all, 2 = slightly, 3 = somewhat, 4 = very, and 5 = extremely.

Table 21 displays that across all school poverty levels (low, mid, and high), the highest mean responses were seen on *philosophy of teaching and learning* (4.00, 3.91, 3.88) and *desire to promote student understanding* (3.93, 3.99, 3.94). Lowest means were seen on *parents* (1.44, 1.44, 1.40) and *student disruptive behavior* (1.48, 1.56, 1.62). Across all school poverty levels (low, mid, and high), highest standard deviations were seen in *desire for student success* (1.23, 1.24, 1.30), *student motivation and engagement* (1.15, 1.16, 1.13), and *student absenteeism* (1.11, 1.10, 1.15). Lowest standard deviations were seen in *parents* (.67, .75, .70) and *student disruptive behavior* (.78, .76, .73). Additionally, responses to the item *formal or informal school or district policies* displayed a notably lower standard deviation (.89) in low-poverty schools as compared to mid- (1.06) and high-poverty (1.08) schools.

The low-poverty school group displayed comparatively high between-group means for *philosophy of teaching and learning* (4.00), while the high-poverty school group displayed notably high between-group means for *desire to accommodate student individual differences and needs* (3.54), *student motivation and engagement* (3.30), *school administrators* (1.90), and *student absenteeism* (2.56). The low-poverty school group displayed comparably low between-group

Table 21

Means, Standard Deviations, and Medians Comparing Influences of Grading by School Poverty Group

| Influences of Grading | School Poverty Level | | | | | | | | |
|---|----------------------|-----------|------------|----------|-----------|------------|----------|-----------|------------|
| | Low | | | Mid | | | High | | |
| | <i>M</i> | <i>SD</i> | <i>Mdn</i> | <i>M</i> | <i>SD</i> | <i>Mdn</i> | <i>M</i> | <i>SD</i> | <i>Mdn</i> |
| 1) Desire for Student Success | 3.33 | 1.23 | 3.00 | 3.29 | 1.24 | 3.50 | 3.32 | 1.30 | 3.00 |
| 2) Philosophy of Teaching and Learning | 4.00 | .96 | 4.00 | 3.91 | .93 | 4.00 | 3.88 | .95 | 4.00 |
| 3) Desire to Promote Student Understanding | 3.93 | .96 | 4.00 | 3.99 | .91 | 4.00 | 3.94 | 1.00 | 4.00 |
| 4) Desire to Accommodate Student Individual Differences and Needs | 3.19 | 1.00 | 3.00 | 3.31 | .97 | 3.00 | 3.54 | 1.02 | 4.00 |
| 5) Student Motivation and Engagement | 3.01 | 1.15 | 3.00 | 3.07 | 1.16 | 3.00 | 3.30 | 1.13 | 3.00 |
| 6) State Standardized Testing | 1.73 | .99 | 1.00 | 1.74 | 1.12 | 1.00 | 1.63 | .99 | 1.00 |
| 7) Formal or Informal School or District Policies | 1.95 | .89 | 2.00 | 2.15 | 1.06 | 2.00 | 2.18 | 1.08 | 2.00 |
| 8) School Administrators | 1.70 | .99 | 1.00 | 1.76 | .98 | 1.00 | 1.90 | 1.04 | 2.00 |
| 9) Parents | 1.44 | .67 | 1.00 | 1.44 | .75 | 1.00 | 1.40 | .70 | 1.00 |
| 10) Student Absenteeism | 2.36 | 1.11 | 2.00 | 2.38 | 1.10 | 2.00 | 2.56 | 1.15 | 3.00 |
| 11) Student Disruptive Behavior | 1.48 | .78 | 1.00 | 1.56 | .76 | 1.00 | 1.62 | .73 | 1.00 |
| 12) Differing Student Ability Levels in a Class | 2.30 | .95 | 2.00 | 2.47 | .99 | 3.00 | 2.45 | 1.02 | 2.00 |
| 13) Student Disruptive and/or Non-Supportive Home Environments | 1.73 | .88 | 1.00 | 1.79 | .91 | 2.00 | 1.73 | .87 | 1.00 |

means for *desire to accommodate student individual differences and needs* (3.19), *student motivation and engagement* (3.01), *formal or informal school or district policies* (1.95), *student disruptive behavior* (1.48), and *differing student abilities in a class* (2.30). The high-poverty school group displayed notably low between-group means for *state standardized testing* (1.63).

A MANOVA parametric test was conducted to investigate significant differences between the three school poverty levels on a linear combination of the 13 assessed influences of grading. Assumptions of independence of observations, multivariate normality, and homogeneity of variance/covariance were checked and met (Leech et al., 2011). Pearson product-moment correlations were conducted to test for multicollinearity (Leech et al., 2011). A significant correlation ($r = .613$), which Cohen (1988) considered a medium to high effect size, existed between Influences 7 and 8. To address this possible source of multicollinearity, Influence 7 was eliminated as a dependent variable from the MANOVA test.

The MANOVA tests found no statistically significant differences between the three school poverty levels, as shown in Table 22. Despite this lack of statistical significance, follow-up univariate ANOVAs were conducted to determine whether the school poverty groups differed on each individual influence on teacher grading rationale. No significant differences were found between groups. However, Influence 4: *desire to accommodate student individual differences and needs* displayed a low p -value (.076) near the .05 significance level. Table 23 displays the complete results.

Qualitative Data Analysis

Qualitative data were used to investigate the rationale used by teachers of different school poverty levels to make grading decisions, and qualitative data were used to triangulate quantitative results on teacher grading practices and the

Table 22

Multivariate Tests for Influences of Grading by School Poverty Level

| Test | Value | F | Df | Error df | p | Partial ETA Squared |
|-----------------------|-------|------|-------|----------|------|---------------------------|
| Pillai's Trace | .080 | .82 | 24.00 | 476.000 | .710 | .040 |
| Wilks' Lambda | .921 | .83 | 24.00 | 474.000 | .702 | .040 |
| Hotelling's Trace | .085 | .83 | 24.00 | 472.000 | .693 | .041 |
| Roy's Largest Root | .075 | 1.49 | 12.00 | 238.000 | .130 | .070 |

influences of teacher grading by school poverty level. Data were collected with focus groups and the Web-based survey questionnaire.

Focus groups provided the primary source of qualitative data. A total of four focus groups were conducted: one took place in a low-poverty school, one in a mid-poverty school, and two took place in high-poverty schools. A description of the four focus groups is displayed in Table 24. Each focus group consisted of two to five regular education teachers of English, foreign language, mathematics, science, or social studies, and they ranged in length from 43 to 57 minutes. Focus group transcripts, which were created by the researcher, resulted in a total of 56 single-spaced pages of data: 15 pages from the low-poverty focus group, 16 pages from the mid-poverty focus group, and 29 pages from the high-poverty focus groups.

As shown in Table 24, the high-poverty school focus groups consisted of a total of 7 teachers and 100 minutes, resulting in the largest amount of qualitative data of the three groups. The low-poverty schools focus group consisted of less than half the number of teachers (3) and approximately half of the time (52 minutes), while the mid-poverty schools focus group consisted of 5 teachers and 47 minutes.

Table 23

Tests of Between-Subjects Effects for Influences of Grading by School Poverty Level

| Test | SS | df | MS | F | p | Partial ETA Squared |
|---|-------|----|-------|-------|------|---------------------------|
| 1) Desire for Student Success | .081 | 2 | .041 | .026 | .975 | .000 |
| 2) Philosophy of Teaching and Learning | .644 | 2 | .322 | .360 | .698 | .003 |
| 3) Desire to Promote Student Understanding | .180 | 2 | .090 | .098 | .906 | .001 |
| 4) Desire to Accommodate Student Individual Differences and Needs | 5.206 | 2 | 2.603 | 2.607 | .076 | .021 |
| 5) Student Motivation and Engagement | 3.792 | 2 | 1.896 | 1.451 | .236 | .012 |
| 6) State Standardized Testing | .633 | 2 | .317 | .295 | .745 | .002 |
| 8) School Administrators | 1.810 | 2 | .905 | .898 | .409 | .007 |
| 9) Parents | .082 | 2 | .041 | .083 | .921 | .001 |
| 10) Student Absenteeism | 2.008 | 2 | 1.004 | .799 | .451 | .006 |
| 11) Student Disruptive Behavior | .783 | 2 | .391 | .687 | .504 | .006 |
| 12) Differing Student Ability Levels in a Class | 1.581 | 2 | .791 | .806 | .448 | .006 |
| 13) Student Disruptive and/or Non-Supportive Home Environments | .228 | 2 | .114 | .145 | .865 | .001 |

Table 24

Description of Focus Group Teacher Make-up

| Focus Group (#) | School Site | School Poverty Level | Teachers (#) | Subjects | Length (minutes) |
|-----------------|-------------|----------------------|--------------|--|------------------|
| 1 | V | High | 5 | English, foreign language, math, science | 57 |
| 2 | N | Mid | 5 | English, foreign language, math, science, social studies | 47 |
| 3 | O | Low | 3 | Foreign language, science, social studies | 52 |
| 4 | M | High | 2 | English, math | 43 |

The secondary source of qualitative data, the survey, helped to provide breadth to the qualitative data. As mentioned in the quantitative section above, all regular education teachers of at least one period of English, foreign language, mathematics, science, or social studies were invited to complete the survey, and these teachers comprised the targeted study population of 915 teachers. A total of 251 teachers completed the survey, but survey completion was narrowly defined as completion of all of the 36 close-ended items used as quantitative data. Items 7 to 36 each included an optional open-ended response to provide additional comments to the close-ended response item. A total of 121 teachers (48.2%) completed at least one open-response item, including 38 low-poverty teachers (46.9%), 40 mid-poverty teachers (46.5%), and 43 high-poverty teachers (51.2%). These data are displayed in Table 25. Completed surveys produced 55 single-

spaced pages of qualitative data: 19 pages from low-poverty schools, 16 pages from mid-poverty schools, and 20 pages from high-poverty schools.

Table 25

Description of Qualitative Survey Participation

| School Site | Participants | % of Total Site Participants |
|--------------------|--------------|------------------------------|
| Low Poverty Level | | |
| I | 12 | 66.7 |
| L | 9 | 52.9 |
| O | 7 | 46.7 |
| Q | 4 | 28.6 |
| W | 6 | 35.3 |
| Total | 38 | 46.9 |
| Mid Poverty Level | | |
| H | 7 | 35 |
| N | 15 | 40.5 |
| P | 3 | 37.5 |
| S | 9 | 90 |
| T | 6 | 54.5 |
| Total | 40 | 46.5 |
| High Poverty Level | | |
| G | 5 | 38.5 |
| J | 4 | 44.4 |
| K | 12 | 75 |
| M | 7 | 70 |
| R | 2 | 50 |
| V | 10 | 47.6 |
| X | 3 | 27.3 |
| Total | 43 | 51.2 |

The constant comparative analysis method employed in qualitative data analysis produced four themes: *promotion of student success*, *grading philosophy and beliefs*, *influences of grading practices*, and *grading practices*. Table 26 displays the themes and qualitative data references by school poverty level. As shown, *grading practices* was referenced most, with 562 occurrences, while *grading philosophy and beliefs* was referenced least, with 154 occurrences. It

should be noted that the high-poverty group had more qualitative data than the two other groups because it consisted of two focus groups. For this reason, the frequency of theme references is not compared by school poverty group.

Table 26

Emergent Themes and Frequency of Data References

| Theme | School Poverty Level | | | Total |
|---------------------------------|----------------------|-----|------|-------|
| | Low | Mid | High | |
| Promotion of Student Success | 110 | 86 | 112 | 308 |
| Grading Philosophy and Beliefs | 37 | 47 | 70 | 154 |
| Influences of Grading Practices | 99 | 112 | 164 | 375 |
| Grading Practices | 190 | 150 | 222 | 562 |

Below, each theme is briefly introduced, and after the results of each theme are presented for each of the three school poverty levels.

Promotion of Student Success

The first emergent theme is *promotion of student success*. This theme began as a category and was identified as the *core phenomenon*—a category around which all other categories were created (Corbin & Strauss, 2008). Because its presence is viewed as central to the human element of the grading process, it is considered an emergent theme separate and equal to the level of the three other themes. *Promotion of student success* refers to teachers' drive and central belief in leading students to succeed in both school and life. The theme was referenced a total of 308 times across all data sources. This theme emerged in teachers' beliefs

and grading practices, as evidenced by subthemes of *philosophy and beliefs* and *practices and policies* (Table 27).

Table 27

Sub-Themes and Categories of the Theme Promotion of Student Success by School Poverty Level

| Sub-Theme | School Poverty Level | | |
|------------------------|--|--|---|
| | Low | Mid | High |
| Philosophy and Beliefs | Emphasis on subjectivity Preparation for the real world Importance of effort Parent-like concern | Emphasis on subjectivity Concerns of over-compensating Importance of effort Parent-like concern | Belief in student ability Importance of effort Integrity of top grades |
| Practices and Policies | Rounding up on borderline grades Class-level dependence Subjectively increasing grades Test retakes and acceptance of late work Emphasis on exceptional circumstances and special needs Emphasis on pass vs. fail | Rounding up on borderline grades Class-level dependence Senior emphasis Minimizing test weight, maximizing classwork weight Test retakes and acceptance of late work Emphasis on exceptional circumstances and special needs Emphasis on pass vs. fail | Rounding up on borderline grades Class-level dependence Minimizing test weight, maximizing classwork weight Test retakes and acceptance of late work Eliminating Ds Emphasis on exceptional circumstances and special needs Emphasis on pass vs. fail |

Low-poverty schools. Teachers in low-poverty schools conveyed their desire for student success, and this was evident in their teaching as well as in their grading philosophy and beliefs. Teachers often emphasized the human element of grading, implying that subjectivity is something important to help students to succeed. One teacher remarked,

So does grading have to be rigid and knowledge and belief-based, and all those other things? Yes, but it also can be just a purely human moment, so we're lucky to be in a profession where you can actually be human.

This statement downplayed the importance of objective considerations of grading and emphasized using grades to help students in a subjective manner. In a similar tone, another teacher declared, "Students are more than just point machines. We need to remember that we are in the people business and all people are not the same." This statement also deemphasized objectivity and uniformity in grading and instead emphasized the "human" and subjective side of grading. A social studies teacher also explained, "I have to prepare them to go out into the world so that they understand these things to make wise decisions later on down the road. That's my guiding force, you know, that's my compulsion." This teacher was among a number of teachers who emphasized their obligation to prepare kids for the "real world," and it was evident that the emphasis impacted their grading practices.

But teachers' desire for student success was not an absolute. Teachers displayed a limit to the lengths they would go to promote success, evidenced by one teacher: "I absolutely want my students to be successful, but my desire for their success can't be greater than their desire (and subsequent effort) for success." This teacher's emphasis of student effort was fairly representative of the entire group, as teachers explained that their effort in promoting student success was often (but not always) contingent upon students' effort in their own success. Additionally, some teachers explained that too much help from teachers would hinder future success, evidenced when a teacher declared, "I will always help those who need it, but I won't hobble their futures [for] them by offering too much assistance." This teacher showed the common concern of teachers of offering too

much support to students, resulting in over-dependent kids unable to succeed on their own.

Teachers also described practices and policies that they used to promote student success such as use of grading curves, extra credit, and especially rounding up on borderline grades. Borderline grades are situations in which a student's percentage grade is on the border between two letter grades—usually slightly below the next letter grade (e.g., 89.2%). This was a common topic, and teachers mostly described their willingness to round to the higher letter grade in these situations, although this rounding was often heavily dependent upon student effort. Some teachers described a straight-forward borderline grade policy, such as “.5 and above are always rounded up.” Others described a more nuanced approach:

When I round, let's say it's a 69.5, I'm going to round it up to a 70, but if it's a 69.3 or .4 and the kid has been trying or I maybe like the person a lot, and they *do* work hard, and it's just maybe test grades, I might round it up to a 70. And I think—are you gonna be nice and give the kid a C instead of a D, or are you going to be mean and keep it as a D?

This teacher's decision-making process included a complex combination of factors including student effort, whether or not she liked the student, and her own disposition in deciding on the borderline grade.

Teachers were especially willing to aid students with grading practices and policies when the borderline grade was between an F and a D or between a D and a C because of the emphasis on students passing a course. In these cases, teachers often described their willingness to pass any and all students if they simply put forth effort. One teacher stated, “Grades are structured so that all students should be able to pass if they continue to turn in missing assignments, I don't fail kids who give their all, at worst they would earn a D.” For this teacher, the meaning of a failing grade was clear: little to no student effort. However, the meaning of all grades above the F seems much less straightforward.

Teacher use of student-friendly practices and policies for students on the border of passing and failing also included emphasizing retakes and late work. One teacher added, “If a student had a high ‘D’ they were given several opportunities to master the content.” This teacher implied a priority of promoting student success for those with lower grades. But teachers also emphasized allowing retakes, late work, and partial credit for all students, as one science teacher explained, “I allow my students to re-take quizzes, resubmit labs for a second grading and I allow them to do missing assignments in my room at lunch for up to full credit depending on the quality of work they do.” The use of test retakes and acceptance of late work to promote student success was the most universal practice that teachers displayed in this theme.

Other teachers described increasing grades in situations other than borderline cases to promote student success: “I am guilty of giving students the ‘benefit of the doubt’ and will give them a higher grade than they sometimes earn. Not proud of that but it's the truth.” This teacher directly acknowledged their bias for students and implied their understanding of the unsoundness of the practice. Another teacher explained the manner in which any student could earn a C: “Effort with revisions of major assignments will allow students to raise grades or earn a C.” For this teacher, like many others in the group, effort is the key factor needed to subjectively raise student grades.

Teachers’ promotion of student success was somewhat contingent upon class level, as students in higher-level classes, including GATE, honors, and Advanced Placement (AP), were not provided with as much teacher grading support as students in college preparatory (CP) and general level classes. One AP teacher remarked, “But also level of class means something too. AP kids—I am

not as forgiving for AP kids.” In the same conversation, another AP teacher replied to this sentiment by commenting,

I don't forgive them. I don't let them slack at all. I offer extra credit to my Spanish 1 classes because they're all freshmen and sophomores mostly, but in my AP class, I don't offer any extra credit because they're already getting extra—whatever they get.

These teachers explained that they do not use grading to promote student success in AP classes like they do in CP and general level classes. The teachers implied that this is a part of the rigor that accompanies the higher-level course, and the latter teacher also explained that the automatic weighting given to AP grades renders additional grading support from teachers unnecessary.

Finally, teachers expressed their willingness to accommodate students with exceptional circumstances and special needs with more lenient grading practices or by making accommodations. One teacher explained, “My students know if there is something going on at home that is impacting their ability to complete work, they just need to let me know and I will make accommodations for them. Unfortunately, they don't always say.” This teacher described students' awareness of their emphasis of promoting student success for students with exceptional circumstances and special needs. Also, the teacher added that because students don't always inform him of their circumstances, the teacher is unable to help with grades as much as he would like.

Mid-poverty schools. Teachers of mid-poverty schools emphasized the importance of student success in their classrooms, and many explained their philosophy and beliefs of pulling for students in their grading practices. One teacher explained, “[I] definitely will error in favor of the student, if there is any question about final grades.” This comment explicitly shows the teacher's bias toward student success in any subjective circumstance. However, teachers were

also concerned about overcompensating and doing students a disservice in the future—especially in higher-level classes. An upper-level teacher stated,

They're going to college. I shouldn't have to worry about giving them a redo for their every quiz that they screw up, you know? I just have to think: What am I teaching my kids? Am I putting in a culture of 'I can screw up and I can redo?' You know, that's my dilemma. Where do I stop?

This teacher's comment reflects a concern expressed in the low-poverty group of providing too much support to produce overly dependent students. Additionally, this teacher emphasized preparing higher-level students for future success in college by utilizing rigorous grading practices and policies.

Teachers in this group largely discussed their practices that are used to promote student success, and often times their philosophy and beliefs were evident through their actions. One teacher explained a grading practice that reinforced this sentiment: "Student success motivates my teaching and re-teaching practices. I may allow one assignment to replace a poorly completed assignment, but their performance earns the grade." This teacher emphasized that grading practices that favor student academic success are the exception, while the rule is objective grading based on student performance. Many teachers went on to explain policies on test retakes and acceptance of late work that aided student academic success, but these common practices (practices common among teachers of the same class or subject) were not as widespread as in the low-poverty group.

Multiple teachers expressed their disappointment that few students take advantage of lenient policies, limiting their effectiveness. One such teacher implied this idea by stating, "I offer students the chance to remediate and retake most evaluation[s]. Those that chose to come in for help generally improve." This teacher implied that student effort to come for extra help from the teacher usually resulted in a higher grade. Many teachers referenced the importance of effort, as

they often explained their push to get students to pass their class with the aid of grading policies and practices, often emphasizing effort. One teacher explained, “I try not to set up my students to fail. They really have to try to do poorly in my class.” This teacher implied that failing in that particular class was reserved for exceptional circumstances.

For senior teachers, graduation was a major concern, and teachers used grades to help students to pass:

I would say with the college prep class, [I consider] how flexible do I need to be to make sure that we hit a reasonable graduation rate, right? How accountable do I hold a kid for—I took an essay today that was 5 weeks late because otherwise that kid would probably, I don’t know, I’m not sure if he would pass or not.

This teacher emphasized that this situation was in a CP class, and he explained his concern for the success of students throughout the school—both at the individual student level and as measured by the school-wide graduation rate.

Teachers explained that they often modified the weight of grading categories (e.g., tests, quizzes, homework, classwork) to maximize student success. One teacher explained her mindset on this practice:

The test grade is what’s killing them all of the time. So then, I get it, some of us don’t do well on tests, I mean, so I don’t want them to flunk out because they freak out on tests, but, um, I have projects embedded in that too, though. So there’s always a chance for them to pass unless they just don’t want to do it at all.

This teacher explained her use of grading policy to promote student success by minimizing the objective test portion of the grade and maximizing more subjective components of the grade. The teacher also added that students have a chance to pass unless they do not put forth any effort. Other teachers specifically mentioned increasing the weight of categories that students traditionally perform best on, such as classwork.

Borderline grades were a common concern for teachers in this group, and teachers often gave students the benefit of the doubt when a report card grade fell between two grades. Teachers were especially lenient in CP and general level classes. A teacher explained, “An F who has a 19% is very different from an F that’s a 59%. So the 59% would certainly get the D. A 50% would probably get a D if it was a college prep kid.” This teacher explained a willingness to potentially lower the D cutoff score to 50%, making the range of scores earning a D from 50% to 69%. The teacher’s willingness to do so seemed to be contingent upon the fact that this was not an upper-level class and that it was a matter of passing versus failing. Many teachers also stated that effort and behavior were a common consideration in these instances. A number explained that although they were not lenient with borderline grades at the quarter, they were flexible at the end of the semester. Teachers explained that they hoped a lower grade at the quarter would serve as motivation for students throughout the rest of the semester, but at the end of the semester, teachers were no longer concerned with motivational purposes of grades.

Mid-poverty teachers explained their willingness to provide exceptions and additional support in exceptional circumstances. A science teacher explained a particular experience:

As a mom, sometimes there’s a kid who you know has had an illness, or a death in the family, or you know, there’s a kid who this year in my class who had to leave the country—his father passed away. He came back, he was behind, and I just dismissed the work during the time. You know, so my personal belief as a mom that the chemistry work that he did was irrelevant for what happened in his life. So in that way my personal beliefs do interfere sometimes, and to the betterment of the kid.

This teacher expressed her concern that exceptional circumstances would get in the way of the academic success of the student, and her role as a mother was a factor in this view and practice.

High-poverty schools. Teachers in high-poverty schools also expressed a strong emphasis on promoting student success. Like teachers of low- and mid-poverty schools, they displayed this emphasis in beliefs and in practices. One teacher explained her belief in inherent student ability by stating, “That idea of, the belief in kids and the belief in their abilities, and the belief in what they can do—it’s either a general policy or it’s one-on-one.” The belief explained by this teacher is clearly one aligned with promoting success, and the teacher also mentioned the use of a grading policy that supports that general belief.

Teachers also emphasized students passing their classes, and they described using their grading practices to help this to happen. One teacher explained, “If they bust it and try, they won't fail. Some gave maximum effort and received D's.” This teacher implied an emphasis on passing students and explained the importance of student effort for the teacher to help with grades. Another teacher reiterated the relative ease of passing a course, but explained that higher grades were reserved for academic achievement: “It is very possible for all students to pass the course but only those who master the content will receive an A or B.” This statement displays the teacher’s emphasis on students passing the class and also on ensuring the integrity of higher grades. Others explained that teacher desire for student success is mitigated by a concern for objective practices: “I desire my students to succeed, so I work really hard to help them do that: tutoring, scaffolding, one-on-one help. I don't just desire success and inflate (or deflate) grades based off something besides my rubrics.” This teacher explained the use of

instructional practices, not grading practices, to help students to succeed, thus helping to ensure objectivity in grading.

As is the case in other poverty levels, borderline grades were a common concern and seen by teachers as an opportunity to promote student success. The primary factor considered in increasing a grade in a borderline situation was effort, as explained by this teacher:

All grades are based on effort. It's a difficult class, even for talented students. However, when considering who to round up and who to not, I will consider their effort. If they did everything I asked them to, asked for help, came in to see me outside of class when they needed it, took every advantage of rewriting and other opportunities to raise their grades, I will honor that effort and round second semester grades within reason.

This teacher provided a fairly extensive definition of effort, including going above and beyond the expectations of all students to have a grade rounded up. Also, as was the case in mid-poverty schools, this teacher mentioned rounding at the semester rather than at the quarter. The teacher did not appear to have a specific rounding policy, as they mentioned rounding “within reason.”

Some teachers discussed using certain practices and policies that increased student grades, although they were often less generous for higher-level classes (GATE, honors, and AP). One teacher explained, “I have built in assignments that are easy to obtain points on to balance out and account for students who do not test well. However, I do not employ different grading scales or percentages for different students unless an IEP requires me to.” As described in low-poverty schools, this teacher modified the types of assignments that were used to create a grade to minimize the negative effect of tests in student grades, effectively increasing student grades without changing the grading system. On the policy side, a math teacher explained that her department eliminated D’s and made it easier for students to earn a C: “In Algebra, our C’s go to 65...but in the GATE,

advanced algebra, it's 70%, because my kids are the ones who are going to go on to honors, math analysis, and calculus." This teacher explained that the common practice in CP and general Algebra was to establish cutoff scores for the C grade at 65%, thus eliminating the D. Additionally, the teacher emphasized utilizing traditional cutoff scores in higher-level classes to prepare students for future success in higher-level mathematics courses.

Finally, like the other school poverty levels, many teachers described policies to improve student grades in the form of test retakes and absent work. There were few exceptions to these policies, and one teacher stated their view of exceptional circumstances:

[I provide the] benefit of the doubt—not on due dates, not always, but a situation, with home stuff going on, or a lot of absences, depending on the kid. There's [*sic*] kids who care about their absences and have no control, and kids who don't care that they're gone all the time, or it's their choice to be gone. But, I have a little more flexibility.

This teacher expressed understanding and flexibility in grading practices and policies in exceptional circumstances—particularly in the case of absences.

Additionally, the teacher expressed some frustration in students who were absent and showed little motivation or effort.

Grading Philosophy and Beliefs

The second emergent theme is *grading philosophy and beliefs*. This theme refers to the meanings and understandings that teachers espouse relating to grades, grading policies, and the grading process. This theme was referenced a total of 154 times across all data sources, and it is composed of five sub-themes: *rigor*, *pedagogical soundness*, *equity and fairness*, *autonomy*, and *meaning of grades* (Table 28).

Table 28

Sub-Themes and Categories of the Theme Grading Philosophy and Beliefs by School Poverty Level

| Sub-Theme | School Poverty Level | | |
|-----------------------|---|--|---|
| | Low | Mid | High |
| Rigor | Strictness Compliance | Meaning of grades | Consistency of standards |
| Pedagogical Soundness | Consistency in practice Emphasis on learning | Students earn grades | Use of percentages |
| Equity and Fairness | Consistency in practice | Students earn grades | Consistency in practice |
| Autonomy | Flexibility for student success Control over practices | Control over practices | -- |
| Meaning of Grades | Frustrations in limitations | Multidimensionality Subjectivity Teacher vs. Student | Frustration in limitations Subjectivity Communicating knowledge |

Low-poverty schools. Teachers in low-poverty schools often displayed a grading philosophy that emphasized rigor. One teacher explained, “Students must submit work that meets the requirements.” This seemingly straightforward statement exemplifies grading theory that emphasizes compliance. On a similar note, another teacher added,

I teach high school, my philosophy is ‘I can teach you but I can't learn you’ so my students sink or swim on their own. Of course I want them to do well but I also think there is value in learning the lesson of failure.

This teacher explained that once the teacher finishes teaching, the learning is up to the student. This teacher’s stated belief in the “lesson of failure,” which

demonstrated a strict policy, implied that he does not believe that grading should be used to promote student success.

Teachers also emphasized the importance of *pedagogical soundness* in their grading practices. To many this term largely included objectivity, and teachers often understood objectivity to be synonymous with consistency, regardless of the grading practice. One social studies teacher explained, “In social studies, we still follow the traditional 90-100 is an A, 80-89 is a B, and you know, again, that really, that’s not a subjective number, that’s, you know, that’s a benchmark.” To this teacher, simply having specific cutoff scores for letter grades was seen as an element of objectivity and rigor in grading. Another teacher explained objectivity in terms of the components of the grade, as the teacher stated, “Grades are based on what you know not how much homework you can do.” This teacher emphasized using grades to communicate learning as opposed to work completion.

The idea of consistency was also important to another core teacher belief: equity and fairness. One particular teacher explained this view along with the objectivity and rigor emphasized in the grading philosophy of this group: “I do not allow negative student behavior to color the way that I grade work (I do not look at the names of the papers as I correct them), but students who choose to disrespect the learning environment will be removed from the class and not be given the opportunity to complete that assignment.” This teacher explained their conscious effort to avoid biases when grading, yet they explained that students misbehaving would not be allowed to display their learning on assignments.

Finally, although some teachers described shared practices, these practices did not seem to constitute a belief. Conversely, a number of teachers expressed a desire for grading autonomy. Teachers described these practices as a type of sacred right of the teacher. One teacher explained, “My grades are my own and

what my students earn. I do not compare grades with other teachers.” This teacher expressed a sense of ownership in grading. Another teacher explained the belief that grading should not only be different for each teacher, but also for each class:

So, you really have to—you really have to know your students, and grading practices are best when designed for the class rather than you know—as close to the teachers desk as possible and as far away from the administrator’s office as possible.

This teacher emphasized grading autonomy to promote student success. These conflicting beliefs regarding grading autonomy display some of the diversity seen in the grading beliefs in the group.

Mid-poverty schools. Many mid-poverty teachers described various meanings of grades used in their classrooms, while others had difficulty in specifying any meaning. One particular teacher alluded to the multidimensionality of grades by commenting, “I know that grades don’t always reflect the learning obviously.” This teacher implied that it is common knowledge that grades often represent meanings other than learning; however, this understanding was far from common knowledge for all teachers in this group. The concerns over the meaning of grades were common, and not all teachers had their own explanation for these meanings. A science teacher described her struggles with the meanings of grades by stating, “I’m gonna give you a C because you are a senior, um, or am I gonna say, when you walk out of physics you gotta know something—that C’s gotta be worth something, or the A’s gotta be worth something.” This teacher explained the temptation to grade subjectively to help her students, but she emphasized the importance of grades representing “something.” Conspicuously, the teacher did not explain what that meaning was or should be.

Furthering the discussion of the meaning of grades, one teacher explained that class level was an important factor in determining meaning:

[Honors students] are graded much more rigorously. So a kid believes that if they turn the assignment in, they deserve an A. I go by the old-fashioned standard an A means excellent, a B means good, a C means average. I'm sorry, that's what it means to me, whether the point value, whatever the point value is, you don't get an A if you turn work in. I expect you to turn your work in. So it is convincing the kids that their work has to be excellent to deserve an A.

As this teacher discussed the rigor in grading higher-level students, she described a frustration in the discrepancy in her students' understanding of the meaning of grades and her own meaning of grades. Additionally, she described grade meanings to be contingent upon quality of completed work. Still others were frustrated with the emphasis on grading. One teacher expressed this frustration by stating, "Education is about learning and improving skills, not grades." This teacher seemed to view grades as different and separate from learning.

Like low-poverty teachers, pedagogical soundness and equity were important ideas in grading to mid-poverty teachers, as they were seen as essential practices of the effective teacher. One teacher explained, "I want my grading scale to be clear and fair, but I don't adjust it to be easy or because I don't believe in failing grades." This teacher seemed to deliberately contrast between grading for fairness and grading to simply help students to receive higher grades. Another teacher expressed her belief in objectivity, although their understanding of the concept in practice seemed unclear: "While assignments might be normalized or curved, grades are objective, so everyone earns their grade." This teacher viewed the subjectivity of curving individual grades as different from the objectivity in creating an overall student report card grade. By stating that students earn their

grades, it seems that objectivity and fairness were important components of grading for this teacher.

Autonomy in grading was also important to mid-poverty teachers. One teacher explained,

I'm the teacher. I can do whatever I want in that class as far as grade cut-offs. If they do their homework all the time, and I know they come for help and they practice on the test and they get a 67%, I can do whatever I want. I can give them a C if I want to.

This teacher emphasized control over his own practices. The teacher was echoed by another who stated, "It is the last true control teachers have in their classroom and the last place that [they] can use their 'art of teaching' skills." These hardline views of control over grading were not necessarily the standard, but they represented at least a sizable opinion in the group.

High-poverty schools. High-poverty teachers often struggled with grade meanings like teachers of the two other poverty levels. One math teacher explained,

I have issues with averaging—like with grades and with tests. Um, I don't know that I can average concepts, you know what I mean. So if tests are over different concepts, um, I can't really weight, you know, solving equations against quadratic functions, and how that—like having one grade to represent all these different concepts—different units—you know—or modules I guess—I think is deceiving. So like a B or a C would mean what? What part of what did you not get? Did you only get like half of everything? Or did you get something like really *really* well and like bombed something else?

This teacher expressed her frustration in the limitations of using a single letter grade to communicate multiple elements of student learning throughout a semester. This frustration seemed to include frustrations with her practices and limitations of improving them. Another teacher expanded on the importance of communicating learning in a grade:

For me, as a teacher receiving students, when I look at a transcript and I see, ‘oh, you got a B in the previous course,’ that needs to mean something to me—that I know that you have above average knowledge of the subject. So when I give a grade, I think it needs to reflect what . . . the level of your working knowledge is.

To this teacher, one important purpose of grades is for communicating the knowledge of students between teachers.

Still other teachers expressed their own frustration in the subjectivity inherent in grading. One teacher explained,

I feel like often times when you’re coming up with that weight on the report card, it’s supposed to be reflective of their learning, but it’s just not. It can be inflated or even deflated by choices that they made or absences that a student has, but I often feel like it’s not a true indicator of what their learning really is.

This particular teacher acknowledged that grades can be subjective and their meaning can be unclear. She also seemed to imply that subjectivity is the fault of the grading system or the student, but not the teacher.

Teachers of high-poverty schools also expressed the importance of pedagogical soundness and equity. One teacher stated,

I try to accommodate learning differences as much as possible, but the place to address that is in classroom instruction, not on grading. Grading should be equally applied as much as possible. A grade in class should be applied equally and consistently and mean exactly the same thing no matter who has that grade.

The teacher explicitly stated the belief that grading should not be used to accommodate student needs or differences. The teacher also used the word “equally” twice in the statement, clearly emphasizing the importance of equity and fairness in grading.

Like teachers of other poverty levels, objectivity was often seen as synonymous with grading practices that used cut-off scores. One teacher explained, “Grades are mathematical after entering the rubric score. A 78 is a 78

and not an 80.” To this teacher, the term “mathematical” is used to express a perceived objectivity because of the use of percentage scores. Additionally, other teachers linked objectivity with the idea of rigor, which teachers felt was essential to the integrity of their teaching. One teacher stated, “I vary how I teach, but I do not lower my standards.” This teacher expressed their rigor in maintaining constant standards.

Influences of Grading Practices

The third emergent theme, *influences of grading practices*, refers to the factors that influence the ways in which teachers conduct their grading practices. This theme was referenced a total of 375 times across all data sources, and it is composed of three sub-themes: *classroom realities*, *pressures*, and *additional considerations* (Table 29).

Table 29

Sub-Themes and Categories of the Theme Influences of Grading Practices by School Poverty Level

| Sub-Theme | School Poverty Level | | |
|---------------------------|---|--|--|
| | Low | Mid | High |
| Classroom Realities | Class level Differing within- and between-class ability levels | Differing within-class ability levels | Class level Differing within-class ability levels |
| Pressures | Administrators | Administrators Students Teachers | Administrators |
| Additional Considerations | Unknown factors | Graduation Integrity and reputation | Graduation Integrity and reputation College Preparation State testing and standards |

Low-poverty schools. Many teachers realized that there were multiple factors that influence grading, but teachers were not always able to describe what those factors exactly were. One teacher explained,

The most difficult [thing] about grading and sitting down and entering grades . . . when you read essays from students, you know, there—if you’re using a rubric, there still is subjectivity involved when you grade that, and so, you have, you know, myriad influences that go into that grade.

This teacher expressed his understanding of the many influences of his grading practices, but by stating “myriad influences,” he seemed to imply that there are many influences that are not known and perhaps cannot be know. Another teacher specified borderline grades as an especially influential grading situation, as the teacher stated, “I’m human. I try to be impartial, but I know sometimes unintentionally those other factors can affect borderline grades.” This teacher also seemed to be unaware of the particular influences impacting grading practices, but the teacher was aware of the subjectivity nonetheless.

Teachers in low-poverty schools explained that student class level—whether general, college preparatory, GATE, honors, or AP—has a significant influence on grading practices. Overall, higher-level classes result in the most rigorous and objective grading practices. One teacher explained, “In my AP course, it is more ability based than my English 9 CP course, which is more completion based.” In this instance, the teacher explained that grades in AP courses were more representative of student learning than in CP courses. However, ability-based learning does not mean entirely objective grades.

Some teachers described difficulties in varying skill levels that exist within the same class, requiring varying grading practices. One teacher stated, “This is a CP assignment and we have no general level classes so the caliber of student I have can range drastically.” This teacher implied that the example assignment was

too difficult for some students of low ability levels and grading was used to accommodate the ability differences. Another teacher explained that the ability levels of entire classes may be so different that different assignments may be needed, which results in different grading. He explained, “There may be a situation where you have a group where they need to do more projects, you know, because that’s just that—every class is different.” In this example, the teacher seemed to imply that classes of lower ability levels require less objective practices. Thus, the teacher implied that projects were more appropriate for the ability level of a class, as opposed to assessments. As a result, grades were likely higher.

Many teachers described feeling different types of pressures on their grading practices. Although some teachers insisted they personally felt no pressure, others described specific experiences with administrators regarding their grading practices. One teacher explained one such experience:

In the past, I’ve had a principal who complained that I was ‘giving too many Fs.’ I pointed out that I could defend every grade I give, and that he was welcome to come spend more time in my class to see what work is and isn’t being completed by the students in the class. He declined my invitation.

The pressure felt by the teacher was accompanied by an antagonism between the two parties. Additionally in this case, the teacher implied that student failures were the result of students not completing assignments. Other teachers explained implicit pressures that were subtler than the previous example:

And there is that reality too, because if you’re giving too many F’s, what happens? ‘Hey, can you come up to the principal’s office today after school? We really need to talk about.’ You know that kind of thing really happens, or, you find that your schedule’s been changed because there’s just too many F’s happening.

In this example, the teacher explained a scenario that he believed would occur if he or another teacher assigned “too many F’s.” This teacher implied that

hypothetical situations such as the one described were an influence on his practices. Also, this teacher did not state that he felt this particular pressure from his own administrators, but rather grading pressure from school administration in general.

Some teachers talked about occasional pressures experienced from parents about grades. One teacher explained, "I think subconsciously that when I know a parent is involved I spend a bit more time on her student's work." These types of pressures seemed to be somewhat infrequent in teacher experiences.

Mid-poverty schools. Like low-poverty teachers, mid-poverty teachers explained numerous influences of their grading practices. One teacher expressed feeling pressure in grading students of differing ability levels within a class. The teacher explained, "I guess in my head, I know that all students should be held to the same standard, but in my heart, it's hard to do that in a run of the mill CP class where students are reading from 3rd grade to PHS." This teacher explained the difficulty remaining objective in grading with the knowledge that certain students have lower ability levels. Thus, the teacher's emotional connection influences grading practices to become more subjective to favor the students' success.

To some teachers of seniors, graduation was a significant influence. One teacher explained,

The consumer math class . . . they will not graduate without these math credits, so when I first taught consumer math course 2 years ago, I had to figure out, OK, so how hard, or how difficult, should I make these tests and exams? Because there is dramatic [difference in] ability levels in these classes.

To this math teacher, differing ability levels coupled with the pressure of student graduation was a significant influence on his grading practices and the rigor of the course.

Other teachers explained that their own integrity and reputation was a significant influence on their grading. One teacher explained,

I'll be honest: even my own external perception [is an influence]. I don't want to be the teacher who all their kids have C's. And I certainly don't want to be the teacher who is in the office for getting spoken to about why my grade book—why kids have bad grades.

This teacher explained that when grading his students, he considered the way in which other parties may perceive his students' grades. The teacher went on to explain that his own comfort level in his practices affected his grading practices: "I feel more comfortable when my kids have good grades as opposed to bad grades—obviously, especially when I'm making an assessment. You like to see A's and B's on a test—I do." In this second statement, the teacher explained that good student grades, especially on assessments, provided less pressure than poor grades.

A number of teachers described pressures from AP students on their grading practices. One teacher stated, "AP calculus, that's you know, the students are very motivated, but they start speaking out if they have a—some students have a 72, or 78—they want the A so bad that when they have a 70 they start freaking out." This teacher explained that by "speaking out," students are unhappy with their grades and apply pressure on the teacher to receive a higher grade.

Some teachers explained that the grades of other teachers have no effect on their own grades. One teacher stated, "I am not usually cognizant of how other teachers grade unless it comes up in a conversation." By being unaware of teacher assigned grades, this teacher implied that he or she is uninfluenced by the grades. However, some teachers explained a low-level of influence of other teachers' grades or the grades of their own students. One teacher stated, "Not that I compare them exactly, but I know what a high level freshmen paper should look

like because of my past grading and norming experiences.” This teacher explained the influence of past student papers as a benchmark for the quality of current students’ papers. Another English teacher explained his concerns of the subjectivity of others’ practices and the impact on higher-level students’ future success:

What if they don’t get into Stanford because their spot was taken by a kid where the teacher you know, say the teacher said 88’s an A, and so, I always wonder, like what is everybody else doing and am I playing the same game as everybody else? Or I am just being obstinate? Is this constructive? How is this really—so I think those are things that I factor in.

This teacher acknowledged the common occurrence of subjective grading in higher-level classes, and he was concerned that his students were put at a disadvantage because of his objective practices. This concern was an influence on his most recent practices—presumably an influence to include more subjectivity into grading.

Like low-poverty teachers, mid-poverty teachers explained feeling some pressure from administration over grading. One teacher explained,

I feel some pressure when I realize that administration is watching over our grades and frowns on the negative opinion that if students are failing, the teacher is doing something wrong. Too many F’s seem to be a red flag. Unfortunately, there can be times when you have a very low performing, disinterested bunch of students. When students don’t do the work, are disengaged, and have no fear of failing, often teacher[s] feel their hands are tied. How can you make them care? This is when the pressure is on. We try many things . . . extra help, tutoring, encouragement, parent contact, extra time on assignments, etc. . . . and yet some students still fail.

This teacher expressed frustration that despite many efforts to help students to succeed, many of the teacher’s students still fail, and as a result, administration provided pressure. The teacher also described feeling continual pressure from administration watching over grades to ensure too many F’s were not assigned.

Another teacher echoed feeling some pressure from administration about grades, but the pressure appears compounded by several additional influences:

My challenge is what can I live with? Cause I have kids that come into physics with what I would consider sixth grade math skills not even there, but they're put in there. Even though I have gone in and said: 'This kid is going to struggle.' But the administration wants them to get a C at least. So what can I live with? Am I gonna say 'You don't even know how to do tangent, sine, cosine, you're in physics, I'm gonna give you a C because you are a senior,' um, or am I gonna say, 'When you walk out of physics you gotta know something—that C's gotta be worth something, or the A's gotta be worth something?'

In this scenario, this teacher described multiple influences of her grading practices: counselors or administrators for enrolling students below requisite ability levels, administrators for applying pressure for students to receive at least a C grade, and her own grading and professional integrity to assign objective grades.

High-poverty schools. Like the other two poverty levels, high-poverty teachers experienced difficulties in their grading because of students of differing skill levels. One teacher explained, "Classes are labeled college prep but range from general to AP honors and include special ed students and EL level 1-4." The great range of abilities and English proficiency levels in this teacher's classes caused difficulties in grading. Inherent in this statement is frustration in these difficulties and the structure of the class.

Teachers also discussed the role of class level on grading practices, along with the pressures that accompany higher-level classes. One teacher explained,

I think another factor would be the level of class you're teaching. If you're teaching a GATE or an AP class, you have a much higher—I think—yoke on your back to pass those kids, versus, a Gen [*sic*] kid can get a D and be OK.

This teacher explained the presence of greater expectations and pressures from outside parties to assign students of higher-level classes at least a C grade.

However, the teacher explained that lower-level students with grades lower than a C would not likely result in additional external pressure.

A number of teachers discussed influences of grading from Smarter Balance (SBAC) standardized testing and other standardized curriculum. One teacher explained,

I would say the pressure too—the SBAC—I know our eleventh grade teachers—the English teachers—they had a pretty positive attitude going into it last year. It's so sketchy—there's so little, I mean there's information, but there's really so little information when you really try to think about.

This particular teacher explained that the pressure was specific to eleventh grade teachers, and it was increased because of the unknown format and content of the test. However, other teachers felt relief from the lack of standardized testing. One teacher stated, “It's just the juniors taking the SBAC, and I don't teach juniors, so I don't really feel a lot of pressure other than to just get them ready for SBAC.”

This teacher implied that eleventh grade teachers do feel pressure on grading because of the standardized testing. However, because she did not teach that level of student, she felt less pressure than in past years in which her students took standardized tests.

Like mid-poverty teachers, some teachers discussed the importance of integrity and reputation in grading. One teacher explained,

And I feel pressured as well too, because I stamp a kid with a C, and they move on. They now have a Smith (pseudonym) stamp with a C for the next teacher of their foundational knowledge should be whatever C level means to that teacher, but it's got my name on it with that kid, and it's like they represent me with that base knowledge, so I'm like ahhh—are you really doing OK, cause she's moving you on, and it's gonna make me look bad. So, I mean there's that factor too, where, don't ruin my reputation, like—just giving everyone C's and just moving them on, so, like, I can't do that.

This math teacher explained the pressure of assigning grades that ultimately decide which students are competent to move to the next course. The teacher explained that she was concerned about her reputation and image as a professional, which in her view would be at least partly measured by the success of her former students.

Senior teachers explained about the pressure of ensuring that students pass their classes and are able to graduate. One teacher stated,

Mainly in May, when I sit there and I'm looking at this, and Tommy is sitting here and he's got a straight 20%, you know it's not that easy when you're making that decision right there: Are you going to graduate or not?

This teacher felt a great amount of pressure close to graduation because her grades were a major factor in determining whether or not students would graduate. She implied that she felt pressure to pass students with any report card grade, no matter how low, so they could graduate. Another teacher explained the pressures of the senior English curriculum:

The senior teachers teach ERWC . . . I know they feel much more of that burden of the C really means something versus—because they need the conditional readiness, they have to get at least C's in that class.

Like the previous senior teacher, this teacher explained influences of grading to give students at least a C in the class so that they are deemed ready for college English.

Finally, like the other poverty groups, teachers felt that administrators added pressure—both direct and indirect—to their grading practices. One teacher explained,

But I think that plays a role a little bit—where you look at a grade and you're like—OK, they're at a 52—they're not at a 22, you know, they're like an F—'Is Henry or Donna (pseudonyms) gonna call me in because I have 8 F's in a CP class. You know, can I go back to see if I didn't grade something correctly, and can they come in to retake it?' So I think the administrator plays a little bit of a role.

This teacher explained some indirect pressure from administrators when it came to assigning F's, because he knew that assigning too many F's would draw attention from administrators. This caused the teacher to provide additional opportunities to failing students to improve their grades.

Grading Practices

The final emergent theme is *grading practices*. This theme is defined as the ways in which teachers create student report card grades. This theme was referenced a total of 562 times across all data sources, and it is composed of five sub-themes: *grading components*, *objective practices*, *subjective practices*, *rigor and procedures*, and *alternative practices* (Table 30).

Table 30

Sub-Themes and Categories of the Theme Grading Practices by School Poverty Level

| Sub-Theme | School Poverty Level | | |
|-----------------------|--|--|--|
| | Low | Mid | High |
| Grading Components | Assessments vs. work Homework | Assessments vs. work Homework | Assessments vs. work Homework |
| Objective Practices | Emphasis of assessments Class-level dependent Rubrics Mastery learning | Emphasis of assessments Class-level dependent | Class-level dependent Rubrics |
| Subjective Practices | Completion of work Assigning of zeros Non-academic factors | Completion of work Assigning of zeros Non-academic factors | Completion of work Assigning of zeros Non-academic factors |
| Rigor and Procedures | Comprehensive grading | Common practices | Common practices |
| Alternative Practices | No D policy Minimum grading | No D policy Minimum grading 4-point grading scale | No D policy Minimum grading |

Low-poverty schools. Teachers often discussed homework and assessments as significant components of their grades, while the relative importance and weight of teachers' grades largely differed. One teacher explained, "Homework is 40% of the grade. Assessment is 60% of the grade." Although this was one of the simpler grading formulas, many teachers explained a similar type of weighting of grading categories. Some teachers emphasized classwork instead of homework, while both classwork and homework were graded either for completeness or correctness. One science teacher explained, "Packets are completion, and they're worth, uh, like 40% of their grade. Something around there." This teacher explained the emphasis that she made in her classroom and her grades on students completing their work, versus emphasizing students learning the material. Other teachers minimized the importance of homework in the grade, instead focusing on assessments. One teacher stated, "While homework is important for students to learn and apply skills on the road to mastery, the burden of the grade should be on mastery assignments: essays, tests, final projects." This teacher emphasized the objective component of the grade, although they conceded that homework was a part of the grade.

A number of teachers explained new grading policies such as a "no D policy" and minimum grading. One teacher explained the use of both of the policies: "We have experimented with a 'No D' policy in which we give a 40% grade for missing homework/classwork. I still give zeros for missing tests, essays and projects." This teacher explained that the "no D policy" was done commonly with other teachers. As previously explained, this policy consists of teachers eliminating the use of the D as a letter grade. Instead, the minimum passing grade is a C. Some teachers increase the range of percentage scores that result in a C, but this teacher did not explain the particular policy. Additionally, the teacher

explained replacing a “0” homework or classwork score in the grade book with a 40%. Interestingly, this teacher explained that this policy did not extend to the most objective measures of student achievement, including tests, essays, and projects. These policies, however, were not necessarily typical, and many teachers stated that they still assigned zeros. One teacher explained, “Students that showed no effort, by not completing and turning in assignments, received a zero on those assignments.” This teacher, like many others, emphasized the importance of effort in a grade.

Additional non-academic grading factors such as effort, participation, behavior, and improvement played varying role in teachers’ grades, but teachers mostly described the role as minimal. One teacher explained the role of effort: “Sometimes you have students that make the effort but are unable to grasp the concept. This student should not be penalized.” This teacher explained that effort was an especially important factor in grades of students of low ability levels. Many teachers explained the indirect effect that the factors played in student grades. One teacher explained, “I don’t grade on participation or paying attention, but in my experience, students who don’t pay attention tend not to do well on assessments.” This common sentiment shows that teachers understood that these non-academic factors do not belong in a grade, but they realized that they played an indirect role in the final grade. Another teacher remarked, “Students that failed, however, did so by choice. No one has ever failed my class because they couldn’t pass, only that they couldn’t be bothered.” This teacher implied that ability level was not ever a factor in their class, but effort was a major factor.

Teachers strived to be objective, and some described the use of rubrics to accomplish this. One teacher explained, “Like I said I read and correct everything.

I use rubrics and I expect quality work.” This teacher emphasized the thoroughness of grading practices and objectivity in grading by using a rubric.

Teachers often described grading practices and policies that were strictest in higher-level classes. One teacher explained, “My AP class is like—it’s due on this day, and you have to have a very good excuse if you’re turning it in late because I might not take it. I might not, I don’t know.” This teacher emphasized the strictness of her policy in her higher-level class, although she appeared to allow room for subjectivity in deciding whether or not to accept work after deadlines.

Mid-poverty schools. Teachers of mid-poverty schools described vastly different methods by which they arrived at report card grades. Like the other groups, assessments and work comprised the two major categories of the grade. One teacher explained, “Homework factors very little into my final grade calculations. However the association between test scores and homework completion is strong.” This teacher implied the importance of assessments in the grade by deemphasizing homework, and they implied the importance of effort by explaining the importance of homework completion to earn high test scores.

Another math teacher explained the rationale behind his homework grades:

I thought that when you make the homework worth a lot, you’re not sure that they actually did—like they copied, parents might have helped them, and then you can actually get the situation where they get a lot of, you know, high homework grades, and then they don’t do well on the tests. They actually pass and not know anything. So, and then if you don’t make homework worth much, they don’t do it, so a happy medium for me is the 15-20% range for homework, 60-70% on the testing, and then I have daily warm-ups, quizzes, stuff like that, maybe in the 5-10% range.

This teacher explained his thought processes involved in finding the proper weight for test and homework categories. Notably, the policy described by this teacher

was determined through experience, not through a previously learned or established teaching philosophy.

While teachers varied between grading homework and classwork for completion or for correctness, most teachers placed an emphasis on assessments. A teacher explained, “The grades were largely based on assessments that I feel reflect the students’ performance level and knowledge.” This emphasis on assessments demonstrates this teacher’s understanding of the importance of objectivity in grading.

Some teachers utilized alternative practices, such as minimum grading and the elimination of the D grade. One teacher explained,

This year in chemistry we tried the experiment of eliminating the D, and I have to tell you my grades have never been higher. There are F’s, no doubt about the fact that there are F’s. There are kids that do not do their work, but if the kids think that they have to have a C, and that you’re not gonna give them a D, they try harder. And then Clarence (pseudonym) and I had a big . . . disagreement on how low to make the C’s and we could not come to consensus at first quarter. So, and he said—well at semester he’s gonna give a few D’s. I ended up giving a few, but kids didn’t think they were gonna get D’s.

This teacher’s explanation of her alternative grading practices displays her satisfaction over the resulting higher grades. This explanation also alludes to the use of common practices, which was often mentioned by teachers. However, even in “common” practices, teachers may apply the practice differently, as described by this teacher. Another teacher explained, “[my] PLC team sets [the] grading scale and rubrics for common assessments.” This teacher provided another example of common practices and a focus on objectivity.

Other teachers explained their use of a 4-point scale in grading, and this was largely seen as a tool to ensure objectivity and fairness. One teacher commented, “I use the four point scale instead of the tradition[al] 100 point scale.

That is an indication of my philosophy and very much determines the fairness of my grades.” The teacher described that the 4-point grading scale, another alternative grading practice, was specifically used to ensure fairness in grading.

While some teachers experimented with minimum grading instead of assigning zeros, many others continued to use zeros. One teacher explained, “If a student fails to complete assigned work, there is a consequence. Most zeros have little impact, but major assignments (projects/reports) are required and do have a significant impact.” This teacher’s use of zeros was a part of their emphasis of strictness and rigor in grading.

Teachers also described grading practices that differed by the class academic level (e.g., general, CP, GATE). One teacher explained,

College prep kids are a little bit different. There is greater flexibility in how you grade them, everything is done with point values, the kids know their grades, they know the expectations, they know what they wanna get, and they either strive for it or they don’t.

In this statement, the teacher seemed to imply that when compared to classes of higher academic levels, CP classes were graded with greater subjectivity.

Additionally, by repeating what CP students “know” regarding these grading practices, the teacher implied that students are comfortable with this type of grading, and that both teacher and student have a similar understanding of the grading practices.

Like teachers of other poverty levels, non-academic factors played a role in grading for teachers of mid-poverty schools, but most teachers described these factors as a small part of their grading. One teacher explained, “Paying attention is a necessary part of success, so while I didn’t BASE a grade on attention, a students [*sic*] who didn’t pay attention did poorly in the class, by implication.” As described by teachers of low-poverty schools, this teacher emphasized not using

subjective factors in grading practices (such as attention), but they described the indirect effect of attention on grades. Another teacher explained her evolving practices on the topic:

Since going through many of the Solution Tree trainings and via discussion with my learning teams, I have not factored behavior/conduct into my grades. That is a separate issue that is focused on through partnership with parents and students.

This teacher's practices became more objective after professional development and discussion with fellow teachers on the topic of behavior. Other teachers described subjective practices that were related to non-academic factors. One teacher explained,

Sometimes if the kid draws attention to himself because of negative behavior, then I probably scrutinize their work a little bit more. And whether that's right or wrong, I know, but that's being honest, if you draw attention to yourself in class, then you know maybe I'm going to look at you a little bit more.

This teacher shed additional light on the impact of non-academic factors in grading, as she described that indirectly, student behavior affected her grading practices even though she knew that it was not a pedagogically sound practice.

High-poverty schools. Teachers in high-poverty schools described various ways of constructing student grades, and like the other poverty groups, assessments and work (i.e., classwork and/or homework) made up the largest grade categories. Some teachers emphasized assessment grades, while others emphasized the opposite. One teacher explained, "About 60% of their final grade is based on reviewable assessments. Kids can take the assessment, and, if they're not pleased, can re-study and re-take the assessment, for as long as they continue to show improvement." This teacher's emphasis on assessments in the grade was also based on mastery learning.

Some teachers wanted to place a bigger emphasis on assessments, but experienced difficulties in doing so. One teacher remarked,

Ideally, the goal is for all students to understand. I would love to put more emphasis in understanding (only 40%), but right now we're focusing on the idea that if students do the work, they will gain understanding.

This teacher explained a philosophy of promoting student understanding, but they felt that certain unmentioned factors did not allow for this to happen. As a result, this teacher settled for equating effort and work completion to understanding and learning. When it comes to work, some teachers graded work based on correctness, while others graded it for completeness. One teacher stated, “Homework was more of nongraded/participation type practice of key concepts.” This is an example of a teacher who used homework entirely for the practice of skills, as presumably, the large majority of the grade was dedicated to the assessing of the skills practiced in homework.

Teachers also explained different grading practices for different academic class levels. One math teacher explained,

CFAs and unit tests measure course objectives and therefore measure student mastery. In algebra, it's 60% assessment, 40% other—which would be everything else. And advanced algebra, GATE, it's 70% assessment, 30% other.

As was the case in the lower school poverty levels, this teacher explained grading practices that were more objective and assessment-driven at higher academic level classes.

Many teachers also described the use of rubrics. One teacher explained her use of a rubric and the inherent challenges:

I tend to use a 3-point rubric, where you know 3 is it's right and you've showed me you've got the work, 2 is you're almost there but you made some small mistakes, you know 1 is you tried, you have something on the paper, you made an effort, 0 is it's blank. So there is some subjectivity in it, and if it's a higher point value, depending on the difficulty and the

involvement of the work in there, it becomes much more subjective as to what it is.

To this math teacher, rubric grading was primarily used as a means of giving partial credit. While it brought a certain amount of objectivity to an otherwise subjective practice of providing partial credit, overall, the practice was seen as more subjective than grading a problem as either right or wrong.

Teachers also described that their grading practices were a part of their common practices. One teacher stated, “And we do give common—common tests, and some of us do actually work together on making sure that we give the same point structure so that again their’s [*sic*] equity across the different teachers.” Although this teacher’s statement describes the use of common practices, it does not appear as if the use of common practices was very extensive.

Like teachers of lower poverty levels, teachers explained that factors such as effort, behavior, improvement, and ability were at times a direct component of grades, while more often they were indirect factors. One teacher remarked,

I have very capable students, but their work effort and ability to persevere through struggle seriously impacts their grades. I tend not to have many As at the end of the semester, and it's not because I don't have students who can earn As, rather that they are unwilling to put the work in for an A and are satisfied with Bs. It is the same with Fs. The students who fail, fail because they were unwilling to do the work, not because the work is too hard.

This teacher emphasized the significance of student effort in the highest and lowest grades. Additionally, the teacher implied that the difference between an A and a B is often about student effort.

Some teachers described alternative practices designed to promote student success. One teacher explained this practice and some difficulties that were a part of the practice:

So if I give you a D—and in math we stopped giving D’s because students who have a D in the class—well they pass and move on, they don’t have

enough knowledge to be successful at the next level. So that's the —that's also the dilemma of—OK, you're at a 68%—you're almost there, but you're not quite, so do I—what do I do?

In this scenario, the teacher described new borderline grading issues that arose because of the alternative grading practice of eliminating D's. However, the teacher explained that the reason for using this practice was not to simply increase student grades, but to ensure that passing students were prepared for the next class in the subject area. Another alternative grading practice used by some teachers was the elimination of zeros, yet other teachers described their continued use of zeros. When asked about the use of zeros in grading, one teacher commented, "Not sure how to answer this. Students who don't complete an assignment or test receive a 0 for that assignment." This teacher appeared confused at the suggestion of assigning something other than a zero for an assignment or test not completed. Clearly, the use of alternative practices is mixed in high-poverty schools.

Summary of Findings

The purpose of this study was to investigate the impact of school poverty level on the teacher grading decision-making process. Quantitative results revealed a number of ways in which school poverty level affected grading practices. Results showed that teachers of low-poverty schools assigned a significantly higher number of A's as compared to mid-poverty schools ($p < .01$, $d = .53$) and high-poverty schools ($p < .01$, $d = .47$). Also, teachers in low-poverty schools assigned significantly fewer F's as compared to mid-poverty schools ($p < .01$, $d = .52$). Additionally, two particular grading practices were shown to have significant differences in their use by school poverty level. The grading practice *quality of student completed homework* was found to be a significantly greater component of teachers' grades in low-poverty schools as compared to teachers in mid-poverty schools ($p = .002$, $r = -.24$) and also in teachers of high-poverty

schools as compared to teachers in mid-poverty schools ($p = .009$, $r = -.20$). The grading practice *student extra credit for academic performance* was found to be a significantly greater component of teachers' grades in low-poverty schools as compared to mid-poverty schools ($p = .006$, $r = -.17$).

Qualitative findings produced four emergent themes across all school poverty levels: *promotion of student success*, *grading philosophy and beliefs*, *influences of grading practices*, and *grading practices*. The first theme, *promotion of student success*, consists of two subthemes: *philosophy and beliefs* and *practices and policies*. The second theme, *grading philosophy and beliefs*, consists of five sub-themes: *rigor*, *pedagogical soundness*, *equity and fairness*, *autonomy*, and *meaning of grades*. The third theme, *influences of grading practices*, consists of three sub-themes: *classroom realities*, *pressures*, and *additional consideration*. The fourth theme, *grading practices*, consists of five sub-themes: *grading components*, *objective practices*, *subjective practices*, *rigor and procedures*, and *alternative practices*. Themes and sub-themes were found to be fairly consistent across all school poverty levels, although a number of nuanced differences were found.

The following chapter discusses and interprets the findings described above by combining quantitative and qualitative data before making conclusions about each research question.

CHAPTER 5: SUMMARY/DISCUSSION/CONCLUSION

Report card grades are extremely important for students' education—both present and future—and to a large extent, their future success in school, the workplace, and even in life (Marzano, 2000). Unfortunately, grades are notoriously subjective, and the meaning of student grades greatly differs from school to school, teacher to teacher, and even student to student (Guskey, 2015). But regardless of the intended meaning of a grade when created by the teacher in the classroom, the assumed meaning of the grade by all interested parties external to that classroom is exclusively one of academic achievement (Allen, 2005; Brookhart, 1991, 1994, 1999, 2009; Cizek et al., 1995/1996; Cross & Frary, 1999; Frary et al., 1993; Guskey, 2015; Marzano, 2000; McMillan, 2001; Randall & Engelhard, 2010; Reeves, 2011; Stiggins et al., 1989; Wormeli, 2006). Thus, when common scenarios occur that rely on student grades such as counselors placing students into appropriate class levels, colleges selecting students for admission, or parents following their child's report cards, inaccurate grades result in poorer educational services at the individual and systemic level.

Even more concerning is the fact that student grades have been shown to be the most subjective in high-poverty schools (Agnew, 1985; Cauley & McMillan, 2000; Howley et al., 2000; McMillan et al., 2002; Office of Educational Research and Improvement, 1994; Zwick & Himmelfarb, 2011). Because student grades have so many purposes, inaccuracy and subjectivity in grades can potentially result in many negative consequences for students with the greatest needs (McMillan, 2001; R. Rothstein, 2004).

The purpose of this transformative explanatory sequential mixed methods study was to understand the impact of school poverty level on the teacher grading

decision-making process. The study took place in a large high school district composed of 18 comprehensive school sites in the South San Joaquin Valley. The target population of the study was 915 regular education teachers of English, foreign language, mathematics, science, and social studies. The primary instrument of the study was a 36-item Web-based survey questionnaire to collect both quantitative and qualitative data to address the first two research questions. The study also utilized an interview protocol to conduct four focus group interviews. Data from the focus groups were combined with qualitative data from the survey to triangulate the quantitative data to address the first two research questions, while the qualitative data were used on its own to address the third research question. Quantitative data were analyzed through the use of descriptive analyses, one-way ANOVAs, Kruskal-Wallis tests, and a MANOVA test, while qualitative data were analyzed with a constant comparative analysis method. Finally, all data were combined to address the guiding research question.

This chapter presents a summary and discussion of findings, implications for policy and practice, recommendations for future research, limitations of the study, and the conclusion of the study.

Summary and Discussion of Findings

Major findings of the study revealed nuanced effects of school poverty level on the teacher grading decision-making process. These effects were more pronounced in the grading practice domain, although even these effects were varied. Effects of school poverty level on the influences of teacher grading and grading rationale were subtle, yet significant.

In this section, quantitative and qualitative data were mixed to address each research question. These findings are first summarized and then discussed for each research question. As the mixed methods followed an explanatory sequential

design, the quantitative results were the primary results used to address Research Questions 1 and 2, and qualitative results were used to further explain the findings in the context of the research questions. Research Question 3 was addressed only with qualitative data.

Research Question 1

The first research question addressed the effect of school poverty level on teacher grading practices. Overall, the findings from the study demonstrated limited, yet varied effects of school poverty level on teacher grading practices. As shown in Table 31, findings included *varied distribution of grades, limited use of objective practices, and significant use of subjective practices*. However, because schools of different poverty levels serve students of vastly different needs (high-poverty schools generally serve students with the highest needs and low-poverty schools generally serve students with the lowest needs), the relatively uniform use of grading practices across school poverty levels likely results in students grades of significantly different purposes and meanings.

Distribution of grades. The distribution of teacher grades varied across school poverty levels, showing that school poverty level did play a role in the practices. Quantitative results showed that teachers of low-poverty schools assigned the greatest proportion of A's and among the lowest proportion of F's. Specifically, results showed that teachers of low-poverty schools assigned significantly more A's than teachers of both mid-poverty and high-poverty schools, while teachers of low-poverty schools assigned significantly fewer F's than teachers of mid-poverty schools. These results differ somewhat from the research, which shows that student grade distributions are fairly uniform across school poverty levels (Randall & Engelhard, 2010; Office of Educational Research and Improvement, 1994). However, a report from the Office of Educational

Table 31

Summary of Findings for Research Question 1

| Finding | School Poverty Level | | |
|---|---|---|---|
| | Low | Mid | High |
| Varied Distribution of Grades | Highest % of A's* F's rounded up | Highest % of F's* F's rounded up | F's rounded up |
| Limited Use of Objective Practices | 23 references in qualitative data Emphasis of assessments Class-level dependent Often misunderstood | 26 references in qualitative data Emphasis of assessments Class-level dependent Often misunderstood | 41 references in qualitative data Class-level dependent Often misunderstood |
| Significant Use of Subjective Practices | 2.10 SGI Most likely to use quality of student completed homework* Most likely to use student extra credit for academic performance* 96 references in qualitative data Completion of work Assigning of zeros Non-academic factors | 2.07 SGI 55 references in qualitative data Completion of work Assigning of zeros Non-academic factors | 2.14 SGI 99 references in qualitative data Completion of work Assigning of zeros Non-academic factors |

Note. *Significant at $p < .01$.

Research and Improvement (1994), a branch of the U.S. Department of Education, found that despite fairly equal grade distributions, eighth grade students from high-poverty schools were somewhat less likely to receive A's on report cards as compared to students from other poverty levels.

Interestingly, despite the differences in assigned A's, teacher assigned F's were not significantly higher in high-poverty schools as compared to low-poverty

schools. Thus, despite the fact that the grading literature shows that students of high-poverty schools have consistently performed lower than students of low-poverty schools on standardized tests that measure academic achievement (Borg et al., 2012; Dahl & Lochner, 2012; Duncan & Murnane, 2014; Heckman, 2006; Reardon, 2011; Sirin, 2005; Stull, 2013), grade distributions between low-poverty and high-poverty schools appear to be fairly similar. A likely explanation for the same reported percentage of assigned F's in both low- and high-poverty schools is the great attention and significance placed on assigning F's. Qualitative data showed that teachers in all poverty levels often explained that students who exhibited effort rarely, if ever, failed a class. Additionally, in cases of borderline grades, teachers were very lenient in rounding up F's to passing grades. This seems to be because of the stigma put on both the student and the teacher when a student fails a class. This pressure serves as a barrier to limit the number of F's assigned by a teacher regardless of school poverty level, student ability level, or student achievement, thus inflating lower grades.

Objective practices. Teachers in all school poverty levels displayed a limited use of objective grading practices. In the survey used in this study, teachers were asked the extent to which they used 17 different grading practices to create their report card grades. Of the 17 practices, only 2 were considered objective and recommended in the grading literature: *use of student academic achievement* and *use of specific learning objectives mastered by students* (Cross & Frary, 1999). This study found that teachers across all school poverty levels responded equally strongly to using these objective practices, collectively reporting they *largely* used these practices.

Findings from the qualitative data provided some support for the quantitative results, but they also displayed a clearer picture of the intricacies of

objective grading practices in the classroom. Teachers in low- and mid- poverty schools emphasized the use of assessments to determine student learning. However, teachers in all three poverty levels reported using objective grading practices more often in higher-level academic classes such as honors, GATE, and AP as opposed to mid- and lower-level academic classes such as CP and general. These findings imply that although teachers in all school poverty levels use objective grading practices, in general and CP level classes, which often comprise the majority of the students at each school site, students are graded with more subjectivity. Grading less objectively and more subjectively results in grades that less accurately reflect student learning and instead reflect a combination of academic achievement and varying combinations of subjective factors.

These types of practices are similar to those first described by Brookhart (1991), and later by others (e.g., Cizek et al., 1996; Cross & Frary, 1999; McMillan, 2001, 2003), as resulting in a *hodgepodge grade*. Brookhart (1991) explained, “A hodgepodge grade of attitude, effort, and achievement, created in an attempt to provide positive feedback to the student about himself or herself, is not the answer” (p. 36). However in this study, teachers across all school poverty levels utilized hodgepodge grading to address concerns in CP and general class levels that purely objective practices would result in significantly low grades. It also stands to reason that in schools in which fewer students are enrolled in higher-level classes, grades are less accurate than in schools in which a greater number of students are enrolled in these higher-level classes. High-poverty schools are more likely to have fewer sections of higher-level classes, and as a result, teachers are more likely to use hodgepodge grading practices.

Subjective practices. Teachers in all school poverty levels displayed a significant use of subjective grading practices. Fifteen of the seventeen grading

practices asked of teachers in the survey were considered subjective practices and were not recommended by researchers as a component of student report card grades (e.g., Cross & Frary, 1999). Subjective grade index (SGI) scores constructed from the survey, which quantified the subjectivity of teacher grading practices, indicated that teachers uniformly utilized fairly low levels of subjectivity, scoring between 2.07 and 2.14. A score of 1.0 was considered maximum objectivity, while a score of 5.0 was considered maximum subjectivity. However, considering the fact that grading experts recommend *no* use of subjective practices (a score of 1.0 on the SGI), the self-reported scores are significant. Additionally, qualitative results showed that teachers often utilized non-academic factors and emphasized completion of homework when creating student grades. Teachers did respond consistently across poverty levels on 13 of the 15 individual subjective practices, and qualitative data produced similar results, indicating a common level of use of subjective practices across school poverty levels.

Subjective practices most used were student ability level, effort, and inclusion of zeros for incomplete assignments or assessments. Teacher scores on these practices ranged from 3.00 to 3.63, indicating the practices were *somewhat* to *largely* a part of their report card grades. Teachers also discussed the use of these three common subjective practices in the qualitative data. Ability level and effort were commonly used to ensure a sense of equity and as a deciding factor in borderline grades. As mentioned above, teachers used subjective practices more often in CP and general level classes, and ability level and effort were two of the these practices they used to increase student grades and minimize the weight of assessments on the grade. However, the inclusion of zeros was a practice most often reserved for higher-level classes, as teachers viewed this subjective practice

as more rigorous in nature. So while grading in higher-level classes tended to be more objective and accurate overall, teachers of these classes chose to use subjective practices, such as the inclusion of zeros, when they resulted in an increase in rigor in the class. Another example of such a practice, although not as common, is including homework (graded for quality) into the grade.

Two subjective grading practices measured by the survey were found to be significantly different in use across school poverty levels: *quality of student completed homework* and *student extra credit for academic performance*. Teachers in low-poverty schools were shown to be significantly more likely than teachers in mid-poverty schools to use both of these subjective practices. Additionally, teachers in high-poverty schools were shown to be significantly more likely than teachers of mid-poverty schools to use quality of student completed homework in grading. Although not found to be significant at the Bonferonni correction level of $p < .017$ (Morgan et al., 2011), teachers in low-poverty schools used extra credit for academic performance much more often than teachers in high-poverty schools (significant at $p < .05$). Because any differences in subjective practices across school poverty levels were inconsistent, it is difficult to make any conclusions about the impact of poverty. Despite this inconclusiveness, the existence of some impact of school poverty level on teacher subjective grading practices is apparent.

Research Question 2

The second research question addressed the effect of school poverty level on influences of teacher grading. The findings demonstrated that school poverty level had little direct effect on these influences. Of the 13 different influences measured in the survey, none measured as significantly different across school poverty levels. Yet as shown in Table 32, teachers across all school poverty levels

did indicate a number of important influences that affected their rationale and grading practices, and findings included *high desire for student success, central philosophy of teaching and learning, existent administrator pressure, and existent influence of absenteeism.*

Table 32

Summary of Findings for Research Question 2

| Finding | School Poverty Level | | |
|---|--|---|---|
| | Low | Mid | High |
| High Desire for Student Success | Mean survey score of 3.33 110 references in qualitative data Importance of effort Emphasis on exceptional circumstances and special needs | Mean survey score of 3.29 86 references in qualitative data Importance of effort Emphasis on exceptional circumstances and special needs | Mean survey score of 3.32 112 references in qualitative data* Importance of effort Emphasis on exceptional circumstances and special needs |
| Central Philosophy of Teaching and Learning | Mean survey score of 4.00 37 references in qualitative data Equity and fairness Rigor Objectivity Autonomy Meaning of grades | Mean survey score of 3.91 47 references in qualitative data Equity and fairness Rigor Objectivity Autonomy Meaning of grades | Mean survey score of 3.88 70 references in qualitative data* Equity and fairness Rigor Objectivity Autonomy Meaning of grades |
| Existent Administrator Pressure | Mean survey score of 1.70 14 references in qualitative data Emphasis on F's | Mean survey score of 1.76 12 references in qualitative data Emphasis on F's | Mean survey score of 1.90 17 references in qualitative data* Emphasis on F's |
| Existent Influence of Absenteeism | Mean survey score of 2.36 9 references in qualitative data Negative effects on grades | Mean survey score of 2.38 12 references in qualitative data Negative effects on grades | Mean survey score of 2.56 14 references in qualitative data* Negative effects on grades |

Note. *Two focus groups were conducted in high-poverty schools, but only one in each of the other two poverty levels.

The relatively equal occurrence of influences of grading across all school poverty levels likely results in varied student grades. This is because each school poverty level serves vastly different types of students with vastly different needs. Because many of the influences of grading impact the teacher's creation of the grade on a case-by-case basis (e.g., student motivation, desire for student success, desire to accommodate student individual differences), it is possible and even likely for a teacher to be cognizant of the general impact of an influence of grades across all students, but unaware of the frequency of occurrence on the individual level. This additive effect is likely much greater in high-poverty schools because of the greater number of students with high needs requiring teacher intervention.

Student success and philosophy. Two findings—*high desire for student success* and *central philosophy of teaching and learning*—largely come from the quantitative results of two influences: *desire for student success* and *philosophy of teaching and learning*. They were interpreted as significant influences of teacher grading, with teacher survey scores of 3.0 (somewhat influential) or higher across all school poverty levels, and they also commonly occurred in qualitative data. The first significant influence, *desire for student success*, registered a mean influential score of 3.31 (between somewhat and very influential) across all poverty groups. This influence was also found in the qualitative data, shown in the emergent theme *promotion of student success* and referenced 308 times. McMillan and Nash (2000) and McMillan (2003) reported a similar emergent category in their studies on teacher grading decision making that became part of the theoretical framework used in this study. They named this category “pulling for students,” and it referred to both grading and assessment practices “that are designed to give students the best opportunity to be successful” (McMillan & Nash, 2000, p. 12).

Teachers in the current study commonly discussed beliefs, practices, and policies that support the finding of a *high desire for student success*. Interestingly, this finding proved to have both objective and subjective components. The objective component included beliefs and practices aimed at increasing rigor to prepare students for future success. Although teachers of all class academic levels (i.e., higher, mid, and lower) addressed this mindset, it was most often put into practice in higher-level classes and with students without significant needs. The subjective component included beliefs and practices aimed at maximizing student success by assigning students the highest possible report card grade. However, teachers often described reserving the application of these beliefs and practices for students, or sometimes even entire classes, that had lower ability levels, scored poorly on tests, exhibited exceptional effort without success, or experienced exceptional personal difficulties.

Despite the fact that teachers across all school poverty levels seemed to adhere to this influence to equal degrees, there is doubtlessly a greater number of students in high-poverty schools with the attributes described by teachers as deserving this type of grade intervention. In this case, grading interventions would produce hodgepodge grading practices. The result is an additive effect in which an increased number of students in high-poverty schools require grading intervention, which results in a significant amount of grade inflation without the teachers being directly aware of the phenomenon.

The next finding, *central philosophy of teaching and learning*, was derived from the influence *philosophy of teaching and learning*, which teachers scored as the highest influence on grading of all 13 influences with a mean score of 3.93, or just below *very influential*. Additionally, qualitative data produced an emergent theme similar to this influence: *grading philosophy and beliefs*, which was

referenced in the data a total of 154 times. The only difference between the two concepts is that *grading philosophy and beliefs* is more specific to grading, as the name implies. The qualitative data showed that *grading philosophy and beliefs* was comprised of subthemes of *rigor, objectivity, equity and fairness, autonomy, and the meaning of grades*.

Teachers' *central philosophy of teaching and learning* provided a mostly objective and rigorous influence on grades that tended to be idealistic. It was idealistic in the sense that this influence would likely serve as the only influence of grading if no other influences existed. However, as many influences exist in reality, teachers have no choice but to lessen the role of their philosophy in their grading practices, and the greater and more significant the additional influences, the less of the role of their own philosophy. This is likely the reason that teachers in schools across SSD could espouse very similar grading philosophies, but produce grades with different meanings. Because of the additive effect of influences of grading, teachers of high-poverty schools are unable to apply their own grading philosophies when creating student grades as often as teachers in low-poverty schools.

Administrators and absenteeism. Two additional findings—*existent administrator pressure* and *existent influence of absenteeism*—were also interpreted as significant despite their low survey scores. The influence *school administrators* received a mean survey score of 1.79 across all groups, indicating a response below the level of *slightly influential*. Additionally, the influence was referenced 43 times within the qualitative data as a category in the sub-theme *pressures* and the theme *influences of grading practices*. Although not found to be significantly different, teachers of high-poverty schools scored this influence highest (1.90) compared to mid-poverty schools (1.76) and low-poverty schools

(1.70). Even though all scores of this influence were relatively low, in this case any score above 1.0 (not at all influential) can be interpreted as significant because of the context and power of this influence. In the qualitative data, teachers in all poverty levels referenced administrator pressure on grading, although it was often indirect pressure. Teachers mostly explained that these pressures were directed toward to the number of D's and F's that they assigned. Because of this, teachers with greater numbers of students who are less inclined for academic success are more likely to experience pressure from administrators. Because high-poverty schools have more of these types of students, it is likely that teachers in these schools are more likely to face pressure from administrators that influences grading as compared to teachers in schools of lower poverty levels.

The influence *student absenteeism* received a mean survey score of 2.43 across all groups, indicating a response between *slightly* and *somewhat influential*. Additionally, the influence was referenced 35 times within the qualitative data as a category within the sub-theme *classroom realities* and the theme *influences of grading practices*. Although *student absenteeism* was not significantly different across groups and it was not scored significantly high among all teachers, it was deemed significant because like the influence *school administrators*, *student absenteeism* is likely to be a significant influence in high-poverty schools. This is because of the greater occurrences of student absences in high-poverty schools (Rogers & Mirra, 2014; R. Rothstein, 2004). Rogers and Mirra (2014) and Richard Rothstein (2004) explained the negative effect of reduced learning time on achievement in students of high-poverty schools that often occurs because of excessive absences. Because student absenteeism is a larger issue in high-poverty schools, teachers in these schools must make decisions about absent students more often, and these decisions are more likely to result in subjective practices such as

the assigning of zeros for missing work or missed assessments. The additive effect of these instances in high-poverty schools adds to the subjectivity and inaccuracy of student grades.

Research Question 3

The third research question investigated the grading rationale used by teachers of the three different school poverty levels to make grading decisions. In key studies on teacher grading decision making, McMillan (2003) and McMillan and Nash (2000) described the difficulty in obtaining results on teacher grading rationale. Consequently, the studies provided little explanation of rationale, other than that it was highly individualized and largely comprised of “a hodgepodge of factors and influences” (McMillan, 2003, p. 38). As described in Chapter 3, one of this study’s goals was to investigate grading rationale for each poverty level by obtaining qualitative results from both survey and focus groups and posing questions specifically designed to elicit teachers’ explanations for their grading practices. As a result, three findings emerged to explain teacher grading rationale: *pedagogical soundness*, *promotion of student success*, and *equity and fairness* (Table 33). These findings were consistent across the three school poverty levels.

Pedagogical soundness. As professional educators, all teachers desire a certain amount of *pedagogical soundness* in their grading practices, and this was found to be the first of three major components of teacher grading rationale, as shown in Figure 3. In this study, *pedagogical soundness* is defined as grading practices deemed by the teacher to be pedagogically strong and congruent with the teacher’s personal philosophy of teaching and learning. It is important to note that *pedagogical soundness* often differed within schools, but there were no significant differences across schools. Additionally, a teacher’s idea of *pedagogical soundness* was most often different from the ideas and research of grading experts.

Table 33

Summary of Findings for Research Question 3

| Finding | School Poverty Level | | |
|------------------------------|---|--|--|
| | Low | Mid | High |
| Pedagogical Soundness | 17 references in qualitative data High standards Professional integrity Resistance to pressures Integrity of the grade Emphasis on objectivity | 18 references in qualitative data High standards Professional integrity Integrity of the grade Emphasis on objectivity | 22 references in qualitative data* High standards Professional integrity Emphasis on objectivity |
| Promotion of Student Success | 110 references in qualitative data Preparation for the future Borderline grades Multiple opportunities | 86 references in qualitative data Preparation for the future Borderline grades Multiple opportunities | 112 references in qualitative data* Preparation for the future Borderline grades Multiple opportunities |
| Equity and Fairness | 21 references in qualitative data Common practices Consistency in practice Meeting needs | 17 references in qualitative data Common practices Consistency in practice Meeting needs | 34 references in qualitative data* Common practices Consideration of ability Meeting needs |

Note. *Two focus groups were conducted in high-poverty schools, but only one in each of the other two poverty levels.

According to teachers within this study, elements including high standards, professional integrity, integrity of grades, and an emphasis on objectivity were important elements of the *pedagogical soundness* finding that helped to explain the rationale that teachers used when grading.

To a large extent, *pedagogical soundness* was closely tied to objective grading practices. Even the most subjective grader desires to establish a certain amount of objectivity in grading, as even he or she inherently understands that objectivity is ultimately desired in grading. Although teachers often

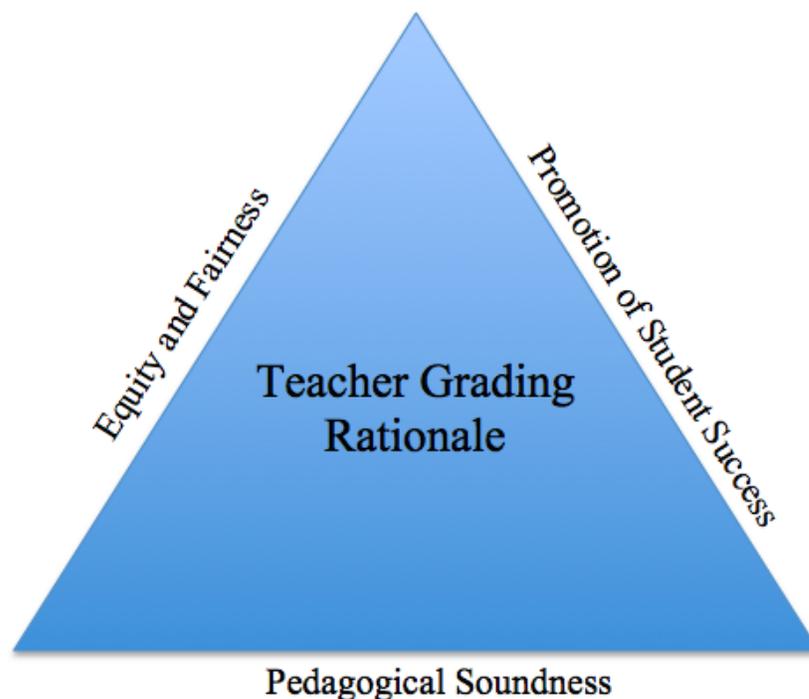


Figure 3. Three significant components of teacher grading rationale.

misunderstand what it means to use objective grading practices, this does not change the fact that teachers value objectivity in grading. Of the three components of teacher rationale presented in this section, *pedagogical soundness* is the only component that a teacher would use independently as a rationale to create student grades. This concept is reflected in Figure 3 by the fact that *pedagogical soundness* represents the base and foundation of the teacher grading rationale triangle. A teacher may reason to be completely pedagogically sound and disregard all other influences and factors to create a student grade. However, this cannot be said for the two other components of grading rationale. Closely related to professionalism, the use of *pedagogical soundness* as a rationale is influenced by a desire to maintain one's integrity and reputation as a teacher. This feeling is also largely variable among teachers.

Finally, the degree to which a teacher employs *pedagogical soundness* as a grading rationale is closely related to the homogeneity of the class. The greater the homogeneity of students, the easier it is for the teacher to grade students in the same way, utilize objective grading practices, allow students to take responsibility for their own success, and ensure fairness. In this manner, students who do not succeed do not have to be aided by the teacher with subjective practices and policies because the lack of success is likely due to a lack of effort. Teachers of all poverty levels expressed little sympathy for students who give little effort.

Promotion of student success. The finding *promotion of student success* is seen as a major component of teacher grading rationale (Figure 3). This concept refers to teachers' intent and predisposition to help students to succeed. Although the concept can apply to many different teaching practices, in this case, it applies to student grades. A similar finding, *high desire for student success*, was discussed in the context of Research Question 2. However, this theme was also seen as a major component of teacher grading rationale. Unlike *desire for student success*, the theme *promotion of student success* is an active role taken by the teacher. Qualitative results showed *promotion of student success* was a very common topic in the data, as teachers emphasized preparing students for the future, rounding up borderline grades, and providing students with multiple opportunities for success.

Unlike *pedagogical soundness*, it is not possible for teachers to only use *promotion of student success* as a rationale for grading practices; they must have a certain amount of pedagogy in mind when conducting grading practices (although it is possible for an individual or an isolated practice to be purely conducted to promote student success). This drive to promote student success is inherent in all

teachers, as it is an inherent drive to wish for the success of those for which one is responsible.

Promotion of student success is more common as a grading rationale when a teacher has students who are in need of support to attain academic success. Otherwise, if students are entirely capable of attaining success on their own, the teacher can rely on *pedagogical soundness*, the foundational rationale, for making grading decisions. When applied, the rationale *promotion of student success* results in subjective grading practices. Because students of mid and low academic class levels (i.e., CP and general levels) are more likely to struggle for success, *promotion of student success* is more likely to be a rationale employed by teachers of these class levels. This rationale may also be common in any other class in which students struggle to attain academic success, such as in classes with students of mixed ability levels or higher-level classes (i.e., GATE, honors, and AP) with students of lower ability levels. Because high-poverty schools are more likely to have students who struggle to attain academic success (R. Rothstein, 2004), these schools are more likely to have teachers who more commonly employ this rationale, and as a result, more often create inaccurate student grades.

Equity and fairness. The third component of teacher grading rationale, as explained by the qualitative results, is *equity and fairness* (Figure 3). Teachers discussed common practices, consistency in practice, and meeting student needs as elements important in grading decisions. Common practices refers to those teaching practices conducted commonly among teachers of the same class or subject, and teachers often discussed their importance in ensuring fairness throughout the school. Consistency in practice refers to teachers' emphasis of applying the same grading practices to all students to ensure equity and fairness.

Meeting student needs refers to teachers using grading practices and policies to ensure equity and fairness in students of high needs.

As was described in the rationale component *promotion of student success*, it is not likely for teachers to only employ an *equity and fairness* rationale for their general grading practices. However, this rationale is extremely important to teachers of all poverty levels, and it is strongly related to both *pedagogical soundness* and *promotion of student success* in its makeup. Unlike either of the other two components of grading rationale, the significant use of *equity and fairness* as a rationale may result in either objective or subjective grading practices. Consistent grading practices may be a result of *equity and fairness* as a rationale, and these practices are likely objective (at least in part). On the other hand, a teacher may wish to increase a low-ability student's grade because his high effort deserves the grade boost. In this case, the teacher may feel that she is ensuring *equity and fairness* of student grades across student ability levels. These types of practices, despite being driven by *equity and fairness*, are largely subjective.

Teachers of students with similar abilities and backgrounds are less likely to apply the *equity and fairness* grading rationale because the teacher would less likely find the need to do so. However, in high-poverty schools, *equity and fairness* is less inherent throughout the school than in low-poverty schools, and teachers in high-poverty schools are more likely to feel compelled to apply the *equity and fairness* grading rationale. As a result, student grades in high-poverty schools are more likely to be subjective and inaccurate than in low-poverty schools.

Guiding Research Question

This study utilized an overarching guiding research question: How does school poverty level affect the teacher grading decision-making process? The teacher grading decision-making process, originally presented and explained in McMillan's (2003) article "Understanding and Improving Teachers' Classroom Assessment Decision Making: Implications for Theory and Practice," is composed of three domains: influences of grading, grading rationale, and grading practices. These three domains culminate in the creation of student grades. The current study produced findings that schools of low-, mid-, and high-poverty levels experienced and utilized similar influences of grading, grading rationale, and grading practices at the systemic level. However, because of poverty's significant impact on students in high-poverty schools, the utilization of the teacher grading decision-making process is different in high-poverty schools than in lower-poverty schools at the student level.

Conceptually, the teacher grading decision-making process was essentially the same across all schools in the study, regardless of school poverty level. When discussing and responding to the process as a whole, teachers revealed few significant differences. However, when grading occurs in the classroom, it is conducted on a student-by-student basis. Although teachers explained that they attempt to be as objective and consistent as possible in their grading practices to maintain pedagogically sound practices that align to their own philosophy of teaching and learning, this is not possible at times. In some instances, teachers emphasized *not* being consistent and objective, such as when grading students of low ability, students displaying high effort but little success, and students of high needs. In other instances they also emphasized *not* being consistent and objective when addressing borderline grades, providing struggling students with multiple

opportunities for success, and dealing with absenteeism and administrative pressure. These instances are all more common in high-poverty schools. High-poverty schools have a much greater number of students that require teachers to deviate from their philosophy and pedagogical soundness rationale with subjective practices. The additive effect that results in high-poverty schools from each of these instances is significant: student grades are more subjective and less accurate than in schools of lower poverty levels.

These findings were used to create Figure 4, which presents a model of the impact of school poverty level on the teacher grading decision-making process. As was the case in the structural components of this study, McMillan's (2003) teacher grading decision-making process was used as the foundation of the model. Displayed in the top of the model is teacher grading in ideal circumstances, which is purely theoretical in the classroom. Grading in ideal circumstances would result in no additional influences of grading, and teacher grading rationale would only be based on pedagogical soundness. The result of this rationale would be grading practices that are purely objective, consistent across all students, and completely aligned to a teacher's philosophy of teaching and learning. This theoretical component of the model is important because it displays the idea that teachers seek to grade in pedagogically sound ways that result in objective practices, but circumstances that are impacted by influences on grading cause teachers to deviate from these practices to others that are more subjective.

The bottom half of the model displays the teacher grading decision-making process that occurs in schools with students in poverty. The four influences shown to affect teacher grading rationale—teacher desire for student success, administrator pressure, absenteeism, and teacher philosophy of teaching and learning—are influences that were found to be present across all school poverty levels (see

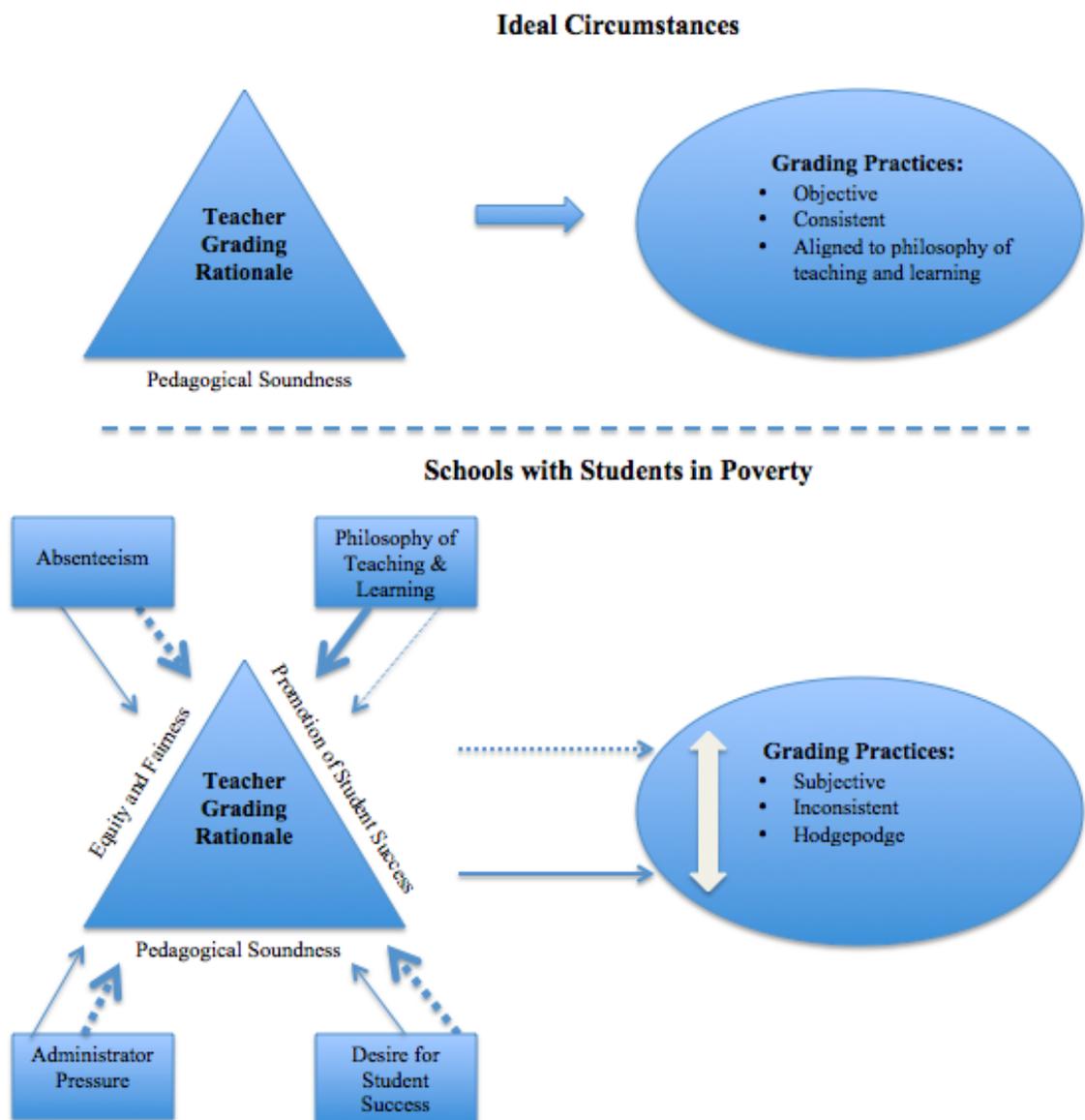


Figure 4. The impact of school poverty level on teacher grading decision making. In the *schools with students in poverty* section of the model, dotted arrows represent high-poverty schools, and solid arrows represent low-poverty schools. Additionally, the weight of each arrow represents the strength of each influence.

discussion of Research Question 2 results). These influences affect teacher grading rationale, which largely consists of pedagogical soundness, equity and fairness, and promotion of student success (see discussion of Research Question 3 results). Finally, this rationale is used by teachers to apply different grading practices, which are subjective, inconsistent, and hodgepodge in nature (see discussion of Research Question 1 results). These practices result in student report card grades that are often subjective and inaccurate.

However, the impact of each influence and the level of subjectivity is affected by school poverty level. In the model, low-poverty schools are depicted by solid arrows and high-poverty schools are depicted by dotted arrows, while the weight of the arrow represents the level of impact of the influence. As shown in the model, student absenteeism, administrator pressure, and teacher desire for student success are great influences on teacher grading rationale in high-poverty schools, while teacher philosophy of teaching and learning is a relatively small influence. Conversely, teacher philosophy of teaching and learning is a great influence on teacher grading rationale in low-poverty schools, while student absenteeism, administrator pressure, and teacher desire for student success are relatively small influences. These influences act on teacher grading rationale to produce high subjectivity, inconsistency, and hodgepodge grading in high-poverty schools, while they result in lower levels of subjectivity, inconsistency, and hodgepodge grading in low-poverty schools.

Implications for Policy and Practice

Grading reform has a long way to go in both policy and practice to transform teacher grading into effective pedagogical practices. Because of the history and tradition involved in grading, change is not easy, and the autonomy in practice that teachers often enjoy presents additional difficulties for systemic

reform. However, when it comes to issues of equity, educators have no room for excuses or delay. This study produced findings that school poverty levels negatively affect student grades, and the higher the poverty level, the greater the negative effect on grades. Because grades play such a large role in the education and future success of students, administrators and teachers have an obligation to ensure that student grades accurately communicate student academic achievement. This obligation applies to the grades of all students, but it is especially crucial and urgent for those students who face among the greatest challenges in their education: students of poverty.

To address this issue, administrators should create school or district grading policies that require all teachers to grade objectively to ensure student grades represent *only* student academic achievement. Student report card grades should be created only from grades that students earn on assessments throughout the marking period. Policy should explicitly state that non-academic factors such as ability, behavior, effort, and improvement should *not* be considered when creating student grades. Additionally, grading policies should require practices such as homework and extra credit to be excluded from grades. Also, schools should clearly establish grade cut-offs so that all letter grades correspond to a pre-established range of percentage scores or grade points (e.g., A = 90-100% or A = 3.7-4.0). In the case of borderline grades, schools should clearly establish rounding procedures.

These changes in policy will require a great amount of support for both teachers and administrators. Teachers and site administrators should be provided with ongoing professional development to educate them about grading theory and effective grading practices. Schools should be prepared for higher than usual failure rates at the outset of the changes, as subjective practices will no longer

inflate student grades. Once the new policies go into effect, it will take a school-wide effort to provide supports and remediation for students as soon as they become in danger of failing.

Of course these recommended changes to policy and practice should occur in all schools, but they are especially important in high-poverty schools. That is why leaders in schools of high poverty must take action on this issue. In these schools, administrators must ensure a substantial amount of support for teachers and students, as the grades of many of these students have likely been inflated since they have been in school. Taking away the subjectivity of grades will likely result in a significant drop in student grades at the outset of the policy change, which could cause substantial negative effects to school morale. But assuming assessments are effective in measuring student achievement, educators, students, and parents should find solace and even satisfaction in the fact that all remediation to improve student grades will directly result in improving student learning deficiencies (something often neglected in common grade-improvement efforts).

In addition to making these changes within schools, schools need to educate the public about the importance of accurate and objective grading practices and the impact of school poverty level on grades. Because the main purpose of grades is to communicate student academic achievement, all interested parties—especially students and parents—should be made aware of this purpose of grading and any changes in policy and practice.

Finally, to prepare future teachers to effectively grade once they enter the classroom, teacher credential programs should require students to complete at least one course that thoroughly addresses both grading theory and effective grading practices. The curriculum of this course should also include decision-making theory and the challenges of grading in schools of high poverty.

Recommendations for Future Research

This study added to the small amount of literature on the impact of poverty on teacher grading practices, but there is a significant need for additional work to be done in this area. While this study utilized a sample from a single school district in the South San Joaquin Valley, a replication of the study utilizing a larger sample throughout the state of California would likely produce more conclusive and significant results that would better address the problem. Results from a study such as this could potentially help to shed additional light on the income achievement gap. Additionally, to better understand teacher grading decision making, further targeted study is needed on the rationale that teachers use to make their grading decisions.

As mentioned in Chapters 1 and 2 of this study, a plethora of research has been conducted documenting poor grading practices and recommendations for best practices, but this has failed to produce any real changes in teacher grading practices. Further study of the grading decision-making process is a way to better understand teacher grading to better facilitate positive reform in this area. This was one of the premises for the current study, and although this study added to the literature in this area, there is a significant need for a study to focus on teacher grading rationale, as this is a complex topic on which data are difficult to collect. Future studies on this topic should consider utilizing psychological theory to better address the topic, as this has not yet been done to the knowledge of this researcher.

Finally, this study addresses a number of influences of teacher grading, and one of the categories was external factors. These factors include, but are not limited to, parents, administrators, state standards, and standardized testing. Because of the possible impact and controversy involving each of these potential factors on teacher grading, a study focused on the importance and impact of the

factors may produce results helpful for administrators who are seeking to create school-wide grading support and policy.

Limitations of the Study

In addition to the limitations described in Chapter 3, a number of study limitations emerged throughout data collection and analysis. Although the original study design called for two focus groups conducted in each school poverty level for a total of six focus groups, scheduling issues at individual school sites resulted in a total of four focus groups: one in the low-poverty school group, one in the mid-poverty school group, and two in the high-poverty school group. Additionally, the size of each focus group differed because of difficulties incurred by site principals in recruiting teacher participation. Instead of all focus groups consisting of five to eight teachers, they consisted of two to five teachers. These limitations reduced the amount of data available for analysis, limiting the potential for significant results. Also, the unequal number and size of focus groups conducted across school poverty levels limited the potential for comparison of qualitative results across groups.

Because the qualitative portion of the survey was optional, not all survey respondents provided qualitative data. The qualitative component of the survey had a sample size of $n = 121$, representing 48.2% of the total sample size of the survey, $n = 251$. Additionally, respondents who provided qualitative data frequently failed to provide data on all 30 open-response items. Thus, although the complete sample size of the survey was composed of 251 respondents who completed all 36 close-ended items of the survey, the sample size of the qualitative portion of the survey was composed of 121 respondents who completed at least one open-ended response (although most respondents included in the qualitative sample size completed multiple open-ended responses). This limitation also

reduced the amount of data available for analysis and limited the potential for additional significant results.

Conclusion

The purpose of this study was to understand the effects of school poverty level on the teacher grading decision-making process. Results, which were interpreted through a critical pedagogy lens, indicated that the direct effects of school poverty level on teacher grading practices, the influences of grading, and teacher grading rationale were intricate and nuanced, yet existent. Although only a small number of quantitative results indicated significant differences between school poverty levels, upon mixing the quantitative and qualitative data, a number of emergent findings were interpreted as significant and causing indirect effects on student grades.

A further interpretation of the findings revealed that teachers likely strive to grade in objective and pedagogically sound ways that are solely influenced by their own philosophy of teaching and learning, but additional influences of grading, including student absenteeism, administrator pressure, and teacher desire for student success, necessitate subjective practices. These latter influences apply more often in students who exhibit exceptional academic or personal needs, and these types of students are more likely to be in poverty. Thus, teachers in high-poverty schools are more likely to experience greater influences of student absenteeism, administrator pressure, and desire for student success. These influences trigger a reaction in teachers' grading rationale to promote student success and ensure equity and fairness, in addition to their foundational rationale of pedagogical soundness, which results in grades that are highly subjective and inaccurate. Because teachers of low-poverty schools rarely have students with exceptional needs, they are often free from the influences on their rationale that

lead to subjectivity, and they are able to grade in an objective and accurate manner.

Subjective and inaccurate grading practices mean student grades do not represent students' true level of academic achievement. Therefore, there is no way for schools and parents to know students' level of learning. The implications of this are great: There is no way to know which students need remediation because of low levels of learning or which students need celebration because of high levels of learning. Alarming, these issues seem to take place most in high-poverty schools—places in which student *most* need remediation for low levels of learning and celebration for high levels of learning.

Perhaps most importantly, this study aimed to better the lives of a marginalized group: impoverished students. Improving grading practices may not immediately impact the lives of these students, but it does have the possibility of indirectly doing so by improving their education. Thus I call on teachers and administrators—across all school poverty levels, but especially of high-poverty schools—to reject the false generosity displayed by providing students with subjective and inaccurate grades and boldly move forward to ensure that grades are objective, accurate, and only representative of student academic achievement. Paolo Freire (1996) aptly explained,

True generosity consists precisely in fighting to destroy the causes which nourish false charity. False charity constrains the fearful and subdued, the 'rejects of life,' to extend their trembling hands. True generosity lies in striving so that these hands—whether of individuals or entire peoples—need be extended less and less in supplication, so that more and more they become human hands which work and, working, transform the world. (p. 45)

REFERENCES

- Agnew, J. E. (1985). *The grading policies and practices of high school teachers*. Paper presented at the meeting of the American Educational Research Association, Chicago.
- Airasian, P., & Jones, A. (1993). The teacher as applied measurer: Realities of classroom measurement and assessment. *Applied Measurement in Education*, 6(3), 241-254.
- Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2001). Schools, achievement, and inequality: A seasonal perspective. *Educational Evaluation & Policy Analysis*, 23(2), 171-191.
- Allen, J. (2005). Grades as valid measures of academic achievement of classroom learning. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 78(5), 218-223.
- Apple, M. (2000). *Official knowledge: Democratic education in a conservative age*. New York, NY: Routledge.
- Ary, D., Jacobs, L., Razavieh, A., & Sorensen, C. (2006). *Introduction to research in education*. Belmont, CA: Thomson.
- Austin, S., & McCann, R. (1992, March). "Here's another arbitrary grade for your collection": A Statewide study of grading policies. Paper presented at the annual meeting of the American Educational Research Association, San Francisco. (ERIC Document Reproduction Service No. ED343944)
- Ayer, F. C. (1933). Chapter Three: School Marks. *Review of Educational Research*, 3(3), 201-204.
- Azeem, M., Afzal, M. T., & Majoka, M. I. (2010). How and why grading. *The International Journal of Learning*, 17(3), 579-596.

- Banks, J. A., & Banks, C. A. (1997). *Multicultural education: Issues and perspectives* (3rd ed.). Boston: Allyn and Bacon.
- Barnes, S. (1985). A study of classroom pupil evaluation: The missing link in teacher education. *Journal of Teacher Education*, 36(4), 46-49.
- Beacon Economics. (2014). South San Joaquin Valley: A growing California success story. Los Angeles: Beacon Economics.
- Best, J. W., & Kahn, J. V. (2006). *Research in education* (10th ed.). Boston, MA: Pearson.
- Bishop, J. H. (1992). Why U.S. students need incentives to learn. *Educational Leadership*, 49(6), 15-18.
- Bomer, R., Dworin, J., May, L., & Semingson, P. (2008). Miseducating teachers about the poor: A critical analysis of Ruby Payne's claims about poverty. *Teachers College Record*, 110(12), 2497-2531.
- Borg, J., Borg, M., & Stranahan, H. (2012). Closing the achievement gap between high-poverty schools and low-poverty schools. *Research in Business and Economics Journal*, 5, 1-24.
- Borman, G., & Dowling, M. (2010). Schools and inequality: A multilevel analysis of Coleman's equality of educational opportunity data. *Teachers College Record*, 112(5), 1201-1246.
- Bourdieu, P., & Passeron, J. 1990. *Reproduction in education, society, and culture*. Beverly Hills, CA: Sage.
- Bowers, A.J. (2009). Reconsidering grades as data for decision making: More than just academic knowledge. *Journal of Educational Administration*, 47, 609-629.

- Bowers, A. J. (2011). What's in a grade? The multidimensional nature of what teacher-assigned grades assess in high school. *Educational Research and Evaluation, 17*(3), 141-159.
- Bradburn, N., Sudman, S., & Wansink, B. (2004). *Asking questions: The definitive guide to questionnaire design – For market research, political polls, and social and health questionnaires*. San Francisco, CA: Jossey-Bass.
- Brennan, R. T., Kim, J., Wenz-Gross, M., & Siperstein, G. N. (2001). The relative equitability of high-stakes testing versus teacher assigned grades: An analysis of Massachusetts Comprehensive Assessment System (MCAS). *Harvard Educational Review, 71*(2), 173-216.
- Brookhart, S.M. (1991). Grading practices and validity. *Educational Measurement: Issues and Practice, 10*(1), 35-36.
- Brookhart, S. M. (1993). Teachers' grading practices: Meanings and values. *Journal of Educational Measurement, 30*, 123-142.
- Brookhart, S. M. (1994). Teacher grading: Practice and theory. *Applied Measurement in Education, 7*(4), 279-301.
- Brookhart, S. M. (1999). *The Art and Science of Classroom Assessment: The Missing Part of Pedagogy*. Washington, DC: ERIC Clearinghouse on Higher Education.
- Brookhart, S. M. (2009). *Grading* (2nd ed.). Upper Saddle River, NJ: Pearson.
- Bryman, A. (2006). Integrating quantitative and qualitative research: How is it done? *Qualitative Research, 6*(1), 97-113.
- Bursuck, W. B., Polloway, E. A., Plante, L., Epstein, M. H., Jayanthi, M., & McConeghy, J. (1996). Report card grading and adaptations: A national survey of classroom practices. *Exceptional Children, 62*(4), 301-318.

- Carey, T., & Carifio, J. (2012). The minimum grading controversy: Results of a quantitative study of seven years of grading data from an urban high school. *Educational Researcher, 41*(6), 201-208.
- Carmichael, P. (1997). Who receives federal Title I assistance? Examination of program funding by school poverty rate in New York state. *Educational Evaluation and Policy Analysis, 19*(4), 354-359.
- Cauley, K., & McMillan, J. H. (2000). *Do teachers grade students differently in high poverty middle schools?* Paper presented at the American Educational Research Association, New Orleans, LA.
- Center on the Developing Child at Harvard University. (2010). *The foundations of lifelong health are built in early childhood*. Retrieved from <http://www.developingchild.harvard.edu>
- Cizek, G. J., Fitzgerald, S. M., & Rachor, R. E. (1995/1996). Teachers' assessment practices: Preparation, isolation, and the kitchen sink. *Educational Assessment, 3*(2), 159-179.
- Clark, C., & Peterson, P. (1986). Teachers' thought processes. In Merlin Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 255-296). New York, NY: MacMillan.
- Clinedinst, M. E., Hurley, S. F., & Hawkins, D. A. (2014). *2013 state of college admission*. Alexandria, VA: National Association for College Admission Counseling.
- Coladarci, T. (2006). School size, student achievement, and the "power rating" of poverty: Substantive finding or statistical artifact? *Education Policy Analysis Archives, 14*(28), 1-23.

- Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., et al. (1966a). *Equality of educational opportunity*. Washington, DC: U.S. Government Printing Office.
- Committee on Psychosocial Aspects of Child and Family Health, Committee on Early Childhood, Adoption, and Dependent Care, & Section on Developmental and Behavioral Pediatrics. (2012). Early childhood adversity, toxic stress, and the role of the pediatrician: Translating developmental science into lifelong health. *Pediatrics*, *129*(1), e224-e231.
- Cooper, C. E., & Crosnoe, R. (2007). The engagement in schooling of economically disadvantaged parents and children. *Youth & Society*, *38*(3), 372-391.
- Corbin, J., & Strauss, A. (2008). *Basic of qualitative research: Techniques and procedures for developing grounded theory* (3rd ed.). Los Angeles, CA: Sage.
- Cox, K. B. (2011). Putting classroom grading on the table: A reform in progress. *American Secondary Education*, *40*(1), 67-87.
- Creswell, J. W. (2007). *Qualitative inquiry & research design: Choosing among five approaches* (2nd ed.). Thousand Oaks, CA: Sage.
- Creswell, J. W. (2008). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed.). Thousand Oaks, CA: Sage.
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Thousand Oaks, CA: Sage.
- Creswell, J. W. & Plano Clark, V. W. (2011). *Designing and conducting mixed methods research* (2nd ed.). Thousand Oaks, CA: Sage.
- Crook, S., & Evans, G. (2014). The role of planning skills in the income-achievement gap. *Child Development*, *85*(2), 405-411.

- Cross, L. H., & Frary, R. B. (1999). Hodgepodge grading: Endorsed by students and teachers alike. *Applied Measurement in Education, 12*(1), 53-72.
- Cureton, L. W. (1971). The history of grading practices. *Measurement in Education, 2*(4), 1-8.
- Curren, R. (1995). Coercion and the ethics of grading and testing. *Educational Theory, 45*(4), 425-441.
- Dahl, G., & Lochner, L. (2012). The impact of family income on child achievement: Evidence from the earned income tax credit. *The American Economic Review, 102*(5), 1927-1956.
- Dobbie, W., & Fryer, R. (2011). Are high-quality schools enough to increase achievement among the poor? Evidence from the Harlem Children's Zone. *American Economic Journal: Applied Economics, 3*(3), 158-187.
- Docan, T. N. (1996). Positive and negative incentives in the classroom: An analysis of grading systems and student motivation. *Journal of Scholarship of Teaching and Learning, 6*(2), 21-40.
- Dueck, M. (2014). *Grading smarter, not harder: Assessment strategies that motivate kids and help them learn*. Alexandria, VA: ASCD.
- Duncan, G. J., & Murnane, R. J. (2014). Growing income inequality threatens American education. *Phi Delta Kappan, 95*(6), 8-14.
- Durm, M. W. (1993). An A is not an A is not an A: A history of grading. *The Educational Forum, 57*(Spring), 294-297.
- Erickson, J. A. (2010). Grading practices: The third rail. *Principal Leadership, 10*(7), 22-24.
- Evans, G. W., Brooks-Gunn, J., & Klebanov, P. K. (2011). Stressing out the poor: Chronic physiological stress and the income-achievement gap. *Pathways, (Winter)*, 16-21.

- Evans, G. W., & Rosenbaum, J. (2008). Self-regulation and the income-achievement gap. *Early Childhood Research Quarterly, 23*(4), 504-514.
- Evans-Winters, V., & Cowie, B. (2009). Cross-cultural communication: Implications for social work practice and a departure from Payne. *Journal of Educational Controversy, 4*(1), 1-12.
- Farkas, G., Grobe, R., Sheehan, D., & Shuan, Y. (1990). Cultural resources and school success: Gender, ethnicity, and poverty groups within an urban district. *American Sociological Review, 55*(1), 127-142.
- Feldman, A., Kropf, A., & Alibrandi, M. (1998). Grading with points: The determination of report card grades by high school science teachers. *School Science and Mathematics, 98*(3), 140-148.
- Fetler, M. (1989). School dropout rates, academic performance, size, and poverty: Correlates of educational reform. *Educational Evaluation & Policy Analysis, 11*(2), 109-116.
- Frery, R. B., Cross, L. H., & Weber, L. J. (1993). Testing and grading practices and opinions of secondary teachers of academic subjects: Implications for instructional measurement. *Educational Measurement: Issues and Practice, 12*(3), 23-30.
- Freire, P. (1996). *Pedagogy of the oppressed*. London, UK: Penguin.
- Frisbie, D. A., & Waltman, K. K. (1992). An NCME instructional module on developing a personal grading plan. *Educational Measurement: Issues and Practice, 11*(3), 35-42.
- Gay, L. R., & Airasian, P. (2003). *Educational research: Competencies for analysis and applications*. Columbus, OH: Merrill Prentice Hall.
- Gershenson, S. (2013). Do summer time-use gaps vary by socioeconomic status? *American Educational Research Journal, 50*(6), 1219-1248.

- Gewirtz, S. (1998). Conceptualizing social justice in education: Mapping the territory. *Journal of Education Policy* 13(4), 469-484.
- Gewirtz, S. (2006). Towards a contextualized analysis of social justice in education. *Educational Philosophy and Theory*, 38(1), 69-81.
- Ginsburg, A., Jordan, P., & Chang, H. (2014, August). Absences add up: How school attendance influences student success. Retrieved from <http://www.attendanceworks.org/research/absences-add/>
- Giroux, H. A. (2011). *On critical pedagogy*. New York: Bloomsbury Academic.
- Giroux, H. A. (2004). Critical pedagogy and the postmodern/modern divide: Towards a pedagogy of democratization. *Teacher Education Quarterly*, 31(1), 31-47.
- Giroux, H. A. (1997). *Pedagogy and the politics of hope: Theory, culture, and schooling: A critical reader*. Boulder, CO: Westview Press.
- Glaser, B. G. (1965). The constant comparative method of qualitative analysis. *Social Problems*, 12(4), 436-445.
- Goldwater, O. D., & Nutt, R. L. (1999). Teachers' and students' work-culture variables associated with positive school outcome. *Adolescence*, 34(136), 653-664.
- Gorski, P. C. (2008). Peddling poverty for profit: Elements of oppression in Ruby Payne's framework. *Equity & Excellence in Education*, 41(1), 130-148.
- Graham, P. (2005). Classroom-based assessment: Changing knowledge and practice through preservice teacher education. *Teaching and Teacher Education*, 21(6), 607-621.
- Guskey, T. R. (1994). Making the grade: What benefits students? *Educational Leadership*, 52(2), 15-20.

- Guskey, T. R. (2011). Stability and change in high school grades. *NASSP Bulletin*, 95(2), 85-98.
- Guskey, T. R. (2015). *On your mark: Challenging the conventions of grading and reporting*. Bloomington, IN: Solution Tree Press.
- Guskey, T. R., & Bailey, J. M. (2001). *Developing grading and reporting systems for student learning*. Thousand Oaks, CA: Corwin Press.
- Hackman, H. W. (2005). Five essential components in social justice education. *Equity and Excellence in Education*, 38: 103-109.
- Hair, E., Halle, T., Terry-Humen, E., Lavelle, B., & Calkins, J. (2006). Children's school readiness in the ECLS-K: Predictions to academic, health, and social outcomes in first grade. *Early Childhood Research Quarterly*, 21(4), 431-454.
- Hanna, R., & Linden, L. (2012). Discrimination in grading. *American Economic Journal*, 4(4), 146-168.
- Hanover Research. (2011, February). *Effective grading practices in the middle school and high school environments* (Research Report). Retrieved from http://www.apsva.us/cms/lib2/VA01000586/Centricity/Domain/63/Hanover_Research__Effective_Grading_Practices_in_the_Middle_School_and_High_School_Environments.pdf
- Hanushek, E. A. (1997). Assessing the effects of school resources on student performance: An update. *Educational Evaluation & Policy Analysis*, 19(2), 141-164.
- Hanushek, E., Kain, J., & Rivkin, S. (2004). Disruption versus Tiebout improvement: The costs and benefits of switching schools. *Journal of Public Economics*, 88(9), 1721-1746.

- Hargis, C. H. (1990). *Grades and grading practices: Obstacles to improving education and to helping at-risk students*. Springfield, IL: Charles C. Thomas.
- Hattie, J. (2009). *Visible Learning: A synthesis of over 800 meta-analyses relating to achievement*. New York: Routledge.
- Heckman, J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, 312(5782), 1900-1902.
- Henwood, B., Derejko, K., Couture, J., & Padgett, D. (2015). Maslow and mental health recovery: A comparative study of homeless programs for adults with serious mental illness. *Administration and Policy in Mental Health and Mental Health Services Research*, 42(2), 220-228.
- Hesse-Biber, S. N. (2010). *Mixed method research; Merging theory with practice*. New York, NY: Guilford Press.
- Hills, J. R. (1991). Apathy concerning testing and grading. *Phi Delta Kappan*, 72, 540-545.
- Howard, T. C. (2010). *Why race and culture matter in schools: Closing the achievement gap in America's classrooms*. New York, NY: Teacher's College Press.
- Howley, A., Kusimo, P., & Parrott, L. (2000). Grading and the ethos of effort. *Learning Environments Research*, 3(3), 229-246.
- Hytten, K. (2006). Education for social justice: Provocations and challenges. *Educational Theory*, 56(2), 221-236.
- Hytten, K., & Bettez, S. C. (2011). Understanding education for social justice. *Educational Foundations*, 25(1-2), 7-24.

- Iacus, S. M., & Porro, G. (2011). Teachers' evaluations and students' achievement: a 'deviation from the reference' analysis. *Education Economics, 19*(2), 139-159.
- Jaggia, S., & Tuerck, D. (2000). *Promoting good schools through wise spending*. Unpublished manuscript, The Beacon Hill Institute, Suffolk University.
- Jenlink, P., & Jenlink, L. (2012). Examining leadership as public pedagogy for social justice. *International Journal of Educational Leadership Preparation, 7*(3), 1-16.
- Johnson, J. D., Howley, C. B., & Howley, A. A. (2002, February). *Size, excellence, and equity: A report on Arkansas schools and districts* (pp. 1-33). Athens, OH: Educational Studies Department, College of Education, Ohio University.
- Kelly, S. (2008). What types of students' effort are rewarded with high marks? *Sociology of Education, 81*(1), 32-52.
- Kena, G., Musu-Gillette, L., Robinson, J., Wang, X., Rathbun, A., Zhang, J., . . . Dunlop Velez, E. (2015). *The condition of education 2015* (NCES 2015-144). U.S. Department of Education, National Center for Education Statistics. Washington, DC. Retrieved from <http://nces.ed.gov/pubsearch>
- Kincheloe, J. L. (2004). *Critical pedagogy primer*. New York, NY: Peter Lang.
- Kincheloe, J. L. (2008). *Knowledge and critical pedagogy: An introduction*. Dordrecht: Springer Netherlands.
- Kirschenbaum, H., Simon, S. B., & Napier, R. W. (1971). *Wad-ja-get? The grading game in American education*. New York, Ny: Hart.
- Lee, V. E., & Burkam, D. T. (2002). *Inequality at the starting gate: Social background differences in achievement as children begin school*. Washington, DC: Economic Policy Institute.

- Leech, N. L., Barrett, K. C., & Morgan, G. A. (2011). *IBM SPSS for intermediate statistics: Use and interpretation* (4th ed.). New York, NY: Routledge.
- Lewis, M., & Battle, J. (2002). The increasing significance of class: The relative effects of race and socioeconomic status on academic achievement. *Journal of Poverty*, 6(2), 21-35.
- Liu, G. (2008). Improving Title I funding equity across states, districts, and schools. *Iowa Law Review*, 93(3), 973-1013.
- Lynch, K., & Baker, J. (2005). Equality in education: An equality of condition perspective. *Theory and Research in Education*, 3(2), 131-164.
- Madon, S., Jussim, L., Keiper, S., Eccles, J., Smith, A., & Palumbo, P. (1998). The accuracy and power of sex, social class, and ethnic stereotypes: A naturalistic study in person perception. *Personality and Social Psychology Bulletin*, 24(12), 1304-1318.
- Manzo, A. N., & Burke, J. M. (2012). Increasing response rate in web-based/internet surveys. In L. Gideon (Ed.), *Handbook of survey methodology for the social sciences* (pp. 327-343). New York, NY: Springer.
- Marzano, R. J. (2000). *Transforming classroom grading*. Alexandria, VA: ASCD.
- Maslow, A. H. (1943). A theory of human motivation. *Psychological Review*, 50, 370-396.
- McLaren, P. (1991). Critical pedagogy: Constructing an arch of social dreaming and a doorway to hope. *The Journal of Education*, 173(1), 9-34.
- McLaren, P. (1997). Critical pedagogy. *Teaching Education*, 9(1), 1-1.
- McMillan, J. H. (2001). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice*, 20(1), 20-32.

- McMillan, J. H. (2003). Understanding and improving teachers' classroom assessment decision making: Implications for theory and practice. *Educational Measurement: Issues and Practice*, 22(4), 34-43.
- McMillan, J. H., & Lawson, S. R. (2001). *Secondary science teachers' classroom assessment and grading practices*. Metropolitan Education Research Consortium, Richmond, VA. (ERIC Document Reproduction Service N. Ed 450 158)
- McMillan, J. H., Myran, S., & Workman, D. (2002). Elementary teachers' classroom assessment and grading practices. *The Journal of Educational Research*, 95(4), 203-213.
- McMillan, J. H., & Nash, S. (2000, April). *Teacher classroom assessment and grading practice decision making*. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.
- Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, 18(2), 5-11.
- Meyer, M. (1908). The grading of students. *Science* 28, 243-250.
- Morgan, D. L. (1997). *Focus groups as qualitative research*. Thousand Oaks, CA: Sage.
- Morgan, G. A., Leech, N. L., Gloeckner, G. W., & Barrett, K. C. (2011). *IBM SPSS for introductory statistics: Use and interpretation* (4th ed.). New York, NY: Routledge.
- Moule, J. (2005). Implementing a social justice perspective in teacher education: Invisible burden for faculty of color. *Teacher Education Quarterly*, 32(4), 23-42.

- Murrell, P. (2006). Toward social justice in urban education: A model of collaborative cultural inquiry in urban schools. *Equity & Excellence in Education, 39*(1), 81-90.
- National Scientific Council on the Developing Child. (2005/2014). *Excessive stress disrupts the architecture of the developing brain: Working paper 3*. Updated Edition. <http://www.developingchild.harvard.edu>
- Nieto, S. (1996). *Affirming diversity: The sociopolitical context of multicultural education* (2nd ed.). White Plains, NY: Longman.
- North, C. E. (2006). More than words? Delving into the substantive meaning(s) of “social justice” in education. *Review of Educational Research, 76*(4), 507-535.
- Novak, M. (2000). Defining social justice. *First Things: A Monthly Journal of Religion and Public Life, 108*, 11-13.
- O'Connor, K. (2009). *How to grade for learning, K-12* (3rd ed.). Thousand Oaks, CA: Corwin.
- Office of Educational Research and Improvement. (1994). *What do student grades mean? Differences across schools*. (Office of Research Rep. OR 94-3401). Washington, DC: U.S. Department of Education.
- Office of Educational Research and Improvement. (1992). *Parental satisfaction with schools and the need for standards*. (Office of Research Rep. OR) Washington, DC: U.S. Department of Education.
- Parmigiani, D. (2012). Teachers and decision-making processes: An Italian exploratory study on individual and collaborative decisions. *Canadian Journal of Education, 35*(1), 171-186.
- Payne, R. (1996). *A framework for understanding poverty*. Highlands, TX: Aha! Process.

- Prince, D. L., & Howard, E. M. (2002). Children and their basic needs. *Early Childhood Education Journal*, 30(1), 27-31.
- Puma, M., Karweit, N., Price, C., Ricciuti, A., Thompson, W., & Vaden-Kiernan, M. (1997). *Prospects: Final report on student outcomes*. Report submitted to the U.S. Department of Education Planning and Evaluation Service.
- Randall, J., & Engelhard, G. (2009). Examining teacher grades using Rasch Measurement Theory. *Journal of Educational Measurement*, 46(1), 1-18.
- Randall, J., & Engelhard, G. (2010). Examining the grading practices of teachers. *Teaching and Teacher Education*, 26, 1372-1380.
- Rauschenberg, S. (2014). How consistent are course grades? An examination of differential grading. *Education Policy Analysis Archives*, 22(92), 1-38.
- Ready, D. D. (2010). Socioeconomic disadvantage, school attendance, and early cognitive development: The differential effects of school exposure. *Sociology of Education*, 83(4), 271-286.
- Reardon, S. F. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. In R. Murnane & G. Duncan (Eds.), *Whither opportunity? Rising inequality and the uncertain life chances of low-income children* (pp. 91-116). New York: Russell Sage Foundation Press.
- Reeves, D. B. (2011). *Elements of grading: A guide to effective practice*. Bloomington, IN: Solution Tree Press.
- Resh, N. (2009). Justice in grades allocation: Teachers' perspective. *Social Psychology of Education: An International Journal*, 12(3), 315-325.
- Rogers, J., & Mirra, N. (2014). *It's about time: Learning time and educational opportunity in California high schools*. Los Angeles, CA: UCLA IDEA.

- Roscigno, V., & Ainsworth-Darnell, J. (1999). Race, cultural capital, and educational resources: Persistent inequalities and achievement returns. *Sociology of Education*, 72(3), 158-178.
- Rothstein, J. (2004). College performance predictions and the sat. *Journal of Econometrics*, 121(1/2), 297-317.
- Rothstein, R. (2004). *Class and schools: Using social, economic and educational reform to close the Black-White achievement gap*. New York, NY: Teacher's College Press.
- Rothstein, R. (2008). Whose problem is poverty? *Educational Leadership*, 65(7), 8-13.
- Rothstein, R. (2013). Why our schools are segregated. *Educational Leadership*, 70(8), 50-55.
- Saldana, J. (2012). *The coding manual for qualitative researchers*. Los Angeles, CA: Sage.
- Senk, S., Beckmann, C. , & Thompson, D. (1997). Assessment and grading in high school mathematics classrooms. *Journal for Research in Mathematics Education*, 28(2), 187-215.
- Shavelson, R. (1973). What is *the* basic teaching skill? *Journal of Teacher Education*, 24(2), 144-151.
- Shepard, L. (2008). The role of assessment in a learning culture. *The Journal of Education*, 189(1/2), 95-106.
- Shonkoff, J. P. (2010). Building a new biodevelopmental framework to guide the future of early childhood policy. *Child Development*, 81(1), 357-367.
- Shuffelton, A. (2013). A matter of friendship: Educational interventions into culture and poverty. *Educational Theory*, 63(3), 299-316.

- Simon, N., & Moore Johnson, S. (2015). Teacher turnover in high-poverty schools: What we know and can do. *Teachers College Record*, 117(3), 1-36.
- Starch, D., & Elliott, E. C. (1912). Reliability of the grading of high school work in English. *School Review*, 20, 442-457.
- Starch, D., & Elliott, E. C. (1913a). Reliability of grading work in mathematics. *School Review*, 21, 254-259.
- Starch, D., & Elliott, E. C. (1913b). Reliability of grading work in history. *School Review*, 21, 676-681.
- Shields, C. M. (2004). Dialogic leadership for social justice: Overcoming pathologies of silence. *Educational Administration Quarterly*, 40(1), 111-134.
- Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, 75(3), 417-453.
- Silvernail, D. L., Sloan, J. E., Paul, C. R., Johnson, A. F., & Stump, E. K. (2014). *The relationship between school poverty and achievement in Maine*. Gorham, ME: University of Southern Maine.
- Smallwood, M. L. (1935). *An historical study of examinations and grading systems in early American universities*. Harvard Studies in Education #24. Cambridge, MA: Harvard University Press.
- Smith, J. (2003). Reconsidering reliability in classroom assessment and grading. *Educational Measurement: Issues and Practice*, 22(4), 26-33.
- Smyth, J. (2011). *Critical pedagogy for social justice*. New York, NY: International Continuum Publishing Group.
- Smyth, J. (2012). The *socially just school* and critical pedagogies in communities put at a disadvantage. *Critical Studies in Education*, 53(1), 9-18.

- South, S., Haynie, D., & Bose, S. (2007). Student mobility and school dropout. *Social Science Research, 36*(1), 68-94.
- Stiggins, R. J., Frisbie, D. A., & Griswold, P. A. (1989). Inside high school grading practices: Building a research agenda. *Educational Measurement: Issues and Practice, 8*(2), 5-14.
- Stull, J. C. (2013). Family socioeconomic status, parent expectations, and a child's achievement. *Research in Education, (90)*, 53-67.
- Theoharis, G. (2007). Social justice educational leaders and resistance: Toward a theory of social justice leadership. *Educational Administration Quarterly, 43*(2), 221-258.
- Thorndike, R. M. (2005). *Measurement and evaluation in psychology and education*. Upper Saddle River, NJ: Prentice Hall.
- Thomas, C. N. (2014). Considering the impact of preservice teacher beliefs on future practice. *Intervention in School and Clinic, 49*(4), 230-236.
- Tomlinson, C. A. (2001a). Grading and differentiation: Paradox or good practice? *Theory Into Practice, 44*(3), 262-269.
- Tomlinson, C. A. (2001b). Grading for success. *Educational Leadership, 58*(6), 12-15.
- Toutkoushian, R., & Curtis, T. (2005). Effects of socioeconomic factors on public high school outcomes and rankings. *The Journal of Educational Research, 98*(5), 259-271.
- Tyack, D., & Tobin, W. (1994). The “grammar” of schooling: Why has it been so hard to change? *American Educational Research Journal, 31*(3), 453-479.
- United States Department of Agriculture. (2013, September). *National school lunch program*. Retrieved from <http://www.fns.usda.gov/nslp/national-school-lunch-program-nslp>

- U. S. Department of Education. (2014, April). *Concentration of public school students eligible for free or reduced-price lunch*. Retrieved from http://nces.ed.gov/programs/coe/indicator_clb.asp#info
- Vagias, Wade M. (2006). *Likert-type scale response anchors*. Clemson International Institute for Tourism & Research Development, Department of Parks, Recreation and Tourism Management. Clemson University.
- Van der Klaauw, W. (2008). Breaking the link between poverty and low student achievement: An evaluation of Title I. *Journal of Econometrics*, 142(2), 731.
- Waltman, K. K., & Frisbie, D. A. (1994). Parents' understanding of their children's report card grades. *Applied Measurement in Education*, 7(3), 223-240.
- Webb, D. (2010). Paulo Freire and 'the need for a kind of education in hope'. *Cambridge Journal of Education*, 40(4), 327-339.
- White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin*, 91(3), 461-481.
- Williams, T. T., & Sánchez, B. (2011). Identifying and decreasing barriers to parent involvement for inner-city parents. *Youth & Society*, 45(1), 54-74.
- Willingham, P.L., Pollack, J.M., & Lewis, C. (2002). Grades and test scores: Accounting for observed differences. *Journal of Educational Measurement*, 39(1), 1-37.
- Woodruff, D. J., & Ziomek, R. L. (2004). *High school grade inflation from 1991 to 2003* (Research report series 2004-04). Iowa City, IA: ACT.
- Wormeli, R. (2006). Accountability: Teaching through assessment and feedback, not grading. *American Secondary Education*, 34(3), 14-27.
- Yeh, S. S. (2013). A re-analysis of the effects of KIPP and the Harlem Promise Academies. *Teachers College Record*, 115(4), 1-20.

- Yosso, T. (2005). Whose culture has capital? A critical race theory discussion of community cultural wealth. *Race Ethnicity and Education*, 8(1), 69-91.
- Zhan, M. (2006). Assets, parental expectations and involvement, and children's educational performance. *Children and Youth Services Review*, 28(8), 961-975.
- Zoeckler, L. (2007). Moral aspects of grading: A study of high school English teachers' perceptions. *American Secondary Education*, 35(2), 83-102.
- Zollers, N. J., Albert, L. R., & Cochran-Smith, M. (2000). In pursuit of social justice: Collaborative research and practice in teacher education. *Action in Teacher Education*, 22(2), 1-12.
- Zwick, R., & Green, J. (2007). New perspectives on the correlation of sat scores, high school grades, and socioeconomic factors. *Journal of Educational Measurement*, 44(1), 23-45.
- Zwick, R., & Himelfarb, I. (2011). The effect of high school socioeconomic status on the predictive validity of sat scores and high school grade-point. *Journal of Educational Measurement*, 48(2), 101-121.

APPENDIX A: SURVEY INSTRUMENT

Teacher Grading Decision-Making Survey

Introduction

Thank you for taking the time to open this survey. You are being asked to complete this survey as part of the data collection methods for the doctoral dissertation of Joshua Kunnath through California State University, Fresno. The purpose of this survey is to investigate the grading practices of high school teachers of core subjects from varying school poverty levels. Specifically, the survey will examine grading practices and influences on grading for teachers of English, foreign languages, mathematics, science, and social studies.

Your participation in this survey is completely voluntary and you can withdraw at any time. The survey consists of 36 total questions and will take approximately 10-15 minutes to complete. All responses are completely anonymous and confidential, and all data collected will only be used for the purposes of this independent study or subsequent independent studies. Results of the study will be published in a doctoral dissertation, but no identifying names of the district, schools, administrators, or faculty will be used.

This research study has been reviewed and approved by the Institutional Review Board (IRB) for Studies Involving Human Subjects at California State University, Fresno. We believe there are no known risks associated with this research study; however, as with any online related activity, the risk of a breach of confidentiality is always possible. To the best of our ability your answers in this study will remain confidential. We will minimize any risks by maintaining anonymity of each survey, assigning pseudonyms for each participating school, and saving data in a secure Web-based personal drive.

If you have questions about this project or if you have a research-related problem, you may contact the researcher, Joshua Kunnath, at (586) 337-0385 or at jkunnath@mail.fresnostate.edu, or the dissertation chair, Dr. Mahmoud Suleiman, at (661) 654-3032, or at msuleiman@csub.edu. If you have any questions concerning your rights as a research subject, you may contact the California State University, Fresno Institutional Review Board through Dr. Kenneth Magdaleno at (559) 278-0309 or at kmagdaleno@csubfresno.edu.

Your participation is greatly appreciated, and the successful completion of the study is attributed in part to the time and input that you have provided. For this I thank you.

1. By clicking "I agree" below you are indicating that you are at least 18 years old, have read and understood this consent form, and agree to participate in this research study. Please print a copy of this page for your records.

- I agree
- I do not agree

Teacher Grading Decision-Making Survey**Part 1: Background Information**

2. How many total years have you been teaching?

- Less than 3
- 3 to 6
- 7 to 10
- More than 10

3. What was your major teaching assignment (class taught most often of the five listed below) during the second semester of last school year?

- English
- Foreign Language
- Mathematics
- Science
- Social Studies

4. What proportion of your classes were college preparatory (i.e., C.P., GATE, Honors, A.P.)?

- Half or more of classes college preparatory
- Fewer than half of classes college preparatory

Teacher Grading Decision-Making Survey**Part 2: Grading Practices**

5. What percentage of students were given "A's" as a final second semester grade in your major teaching assignment last school year?

- Less than 3%
- 5 to 10%
- 11 to 20%
- More than 20%

6. What percentage of students were given "F's" as a final second semester grade in your major teaching assignment last school year?

- Less than 3%
- 5 to 10%
- 11 to 20%
- More than 20%

Teacher Grading Decision-Making Survey

Please answer #7-23 below based on your report card grading practices in the second semester of last school year for the major teaching assignment that you indicated in question two. Please use the following 5-point scale below to do so:

(1) not at all (2) slightly (3) somewhat (4) largely (5) completely

7. To what extent were your final second semester student grades in your major teaching assignment based on student ability level?

- (1) not at all
- (2) slightly
- (3) somewhat
- (4) largely
- (5) completely

Comments:

8. To what extent were your final first semester student grades in your major teaching assignment based on student academic achievement?

- (1) not at all
- (2) slightly
- (3) somewhat
- (4) largely
- (5) completely

Comments:

9. To what extent were your final second semester student grades in your major teaching assignment based on student disruptive behavior/conduct?

- (1) not at all
 (2) slightly
 (3) somewhat
 (4) largely
 (5) completely

Comments:

10. To what extent were your final second semester student grades in your major teaching assignment based on student effort?

- (1) not at all
 (2) slightly
 (3) somewhat
 (4) largely
 (5) completely

Comments:

11. To what extent were your final second semester student grades in your major teaching assignment based on student participation and/or paying attention?

- (1) not at all
 (2) slightly
 (3) somewhat
 (4) largely
 (5) completely

Comments:

Teacher Grading Decision-Making Survey

12. To what extent were your final second semester student grades in your major teaching assignment based on student improvement of performance?

- (1) not at all
- (2) slightly
- (3) somewhat
- (4) largely
- (5) completely

Comments:

13. To what extent were your final second semester student grades in your major teaching assignment based on student grade distributions of other teachers?

- (1) not at all
- (2) slightly
- (3) somewhat
- (4) largely
- (5) completely

Comments:

14. To what extent were your final second semester student grades in your major teaching assignment based on student performance compared to other students in your class(es)?

- (1) not at all
 (2) slightly
 (3) somewhat
 (4) largely
 (5) completely

Comments:

15. To what extent were your final second semester student grades in your major teaching assignment based on student performance compared to your students from previous years?

- (1) not at all
 (2) slightly
 (3) somewhat
 (4) largely
 (5) completely

Comments:

16. To what extent were your final second semester student grades in your major teaching assignment based on specific learning objectives mastered by students?

- (1) not at all
 (2) slightly
 (3) somewhat
 (4) largely
 (5) completely

Comments:

Teacher Grading Decision-Making Survey

17. To what extent were your final second semester student grades in your major teaching assignment based on formal or informal school or district policy on the percentage of students who may receive A's, B's, C's, D's, F's?

- (1) not at all
- (2) slightly
- (3) somewhat
- (4) largely
- (5) completely

Comments:

18. To what extent were your final second semester student grades in your major teaching assignment based on student effort, improvement, behavior and/or other non-test indicators for borderline grades?

- (1) not at all
- (2) slightly
- (3) somewhat
- (4) largely
- (5) completely

Comments:

19. To what extent were your final second semester student grades in your major teaching assignment based on student completion of homework (not graded)?

- (1) not at all
 (2) slightly
 (3) somewhat
 (4) largely
 (5) completely

Comments:

20. To what extent were your final second student grades in your major teaching assignment based on quality of student completed homework?

- (1) not at all
 (2) slightly
 (3) somewhat
 (4) largely
 (5) completely

Comments:

21. To what extent were your final second semester student grades in your major teaching assignment based on inclusion of zeros for incomplete assignments or assessments?

- (1) not at all
 (2) slightly
 (3) somewhat
 (4) largely
 (5) completely

Comments:

22. To what extent were your final second semester student grades in your major teaching assignment based on student extra credit for academic performance?

- (1) not at all
- (2) slightly
- (3) somewhat
- (4) largely
- (5) completely

Comments:

23. To what extent were your final second semester student grades in your major teaching assignment based on student extra credit for non-academic performance (e.g., bringing in classroom supplies)?

- (1) not at all
- (2) slightly
- (3) somewhat
- (4) largely
- (5) completely

Comments:

Teacher Grading Decision-Making Survey**Part 3: Factors that Influence Grading**

Please answer #24-36 below based on your report card grading practices in the second semester of last school year for the major teaching assignment that you indicated in question two. Please use the following 5-point scale below to do so:

(1) not at all (2) slightly (3) somewhat (4) very (5) extremely

24. How influential is your desire for your students' success on your grading practices?

- (1) not at all
 (2) slightly
 (3) somewhat
 (4) very
 (5) extremely

Comments:

25. How influential is your philosophy of teaching and learning on your grading practices?

- (1) not at all
 (2) slightly
 (3) somewhat
 (4) very
 (5) extremely

Comments:

26. How influential is your desire to promote student understanding on your grading practices?

- (1) not at all
- (2) slightly
- (3) somewhat
- (4) very
- (5) extremely

Comments:

27. How influential is your desire to accommodate student individual differences and needs on your grading practices?

- (1) not at all
- (2) slightly
- (3) somewhat
- (4) very
- (5) extremely

Comments:

28. How influential is student motivation and engagement on your grading practices?

- (1) not at all
- (2) slightly
- (3) somewhat
- (4) very
- (5) extremely

Comments:

Teacher Grading Decision-Making Survey

29. How influential is state standardized testing on your grading practices?

- (1) not at all
- (2) slightly
- (3) somewhat
- (4) very
- (5) extremely

Comments:

30. How influential are formal or informal school or district policies on your grading practices?

- (1) not at all
- (2) slightly
- (3) somewhat
- (4) very
- (5) extremely

Comments:

31. How influential are your school administrators on your grading practices?

- (1) not at all
- (2) slightly
- (3) somewhat
- (4) very
- (5) extremely

Comments:

32. How influential are parents on your grading practices?

- (1) not at all
- (2) slightly
- (3) somewhat
- (4) very
- (5) extremely

Comments:

Teacher Grading Decision-Making Survey

33. How influential is student absenteeism on your grading practices?

- (1) not at all
- (2) slightly
- (3) somewhat
- (4) very
- (5) extremely

Comments:

34. How influential is student disruptive behavior on your grading practices?

- (1) not at all
- (2) slightly
- (3) somewhat
- (4) very
- (5) extremely

Comments:

35. How influential is differing student ability level in a class on your grading practices?

- (1) not at all
- (2) slightly
- (3) somewhat
- (4) very
- (5) extremely

Comments:

36. How influential is student disruptive and/or non-supportive home environment on your grading practices?

- (1) not at all
- (2) slightly
- (3) somewhat
- (4) very
- (5) extremely

Comments:

APPENDIX B: INTERVIEW PROTOCOL

| | |
|---|-------|
| Date: | Time: |
| Length of Activity: | Site: |
| Participants (gender & subject taught): | |

Introductory Statement

Good afternoon, and thank you for coming. My name is Josh Kunnath. I teach at Highland, and I'm also a doctoral candidate at Fresno State. Does anyone mind if I record the meeting for data collection purposes? This focus group is part of the data collection methods of my dissertation. You may remember a survey on grading practices sent out a couple of months ago. That was also part of my study—it was the first phase. This is the second and final phase, and two other schools were also selected to participate in focus groups.

The purpose of this focus group is to better understand the decision-making process that teachers use to create student report card grades. Like the survey, the focus is on regular education teachers of at least one class of English, foreign languages, mathematics, science, and social studies. Like the survey, your participation is voluntary. Your participation in this focus group will represent your informed consent.

I have a group of questions that I'd like to ask all of you, and I'd like to encourage you to respond openly and honestly. All responses are completely anonymous and confidential, and all data collected will only be used for the purposes of this independent study or subsequent independent studies. I will be recording this session so that I can create a transcript of the session. I will also be taking notes during the session.

Your participation is greatly appreciated, and the successful completion of this study is attributed in part to the time and input that you have provided. For this I thank you.

Focus Group Questions

- 1) What are some challenges you face in creating student report card grades?
- 2) How do you create student report card grades?
- 3) How do your own knowledge, beliefs, and/or values (personal or professional) influence the decisions you make in creating report card grades?
- 4) How do external factors (i.e., state testing, district policy, parents, administrators) influence the decisions you make in creating report card grades?
- 5) How do classroom realities (i.e., social promotion, student absenteeism, disruptive behavior, and differing student ability within classes) influence the decisions you make in creating report card grades?

APPENDIX C: OBSERVATIONAL PROTOCOL

| Focus | Descriptive Notes | Reflective Notes |
|--|--------------------------|-------------------------|
| <ul style="list-style-type: none"> • Physical Setting • Visual Layout | | |
| <ul style="list-style-type: none"> • Description of participants • Description of activities • Description of individuals engaged in activity • Sequence of activity over time • Interactions | | |
| <ul style="list-style-type: none"> • Unplanned events • Concerns | | |
| <ul style="list-style-type: none"> • Participant comments • Emerging codes & themes | | |

Fresno State

Non-Exclusive Distribution License

(to archive your thesis/dissertation electronically via the library's eCollections database)

By submitting this license, you (the author or copyright holder) grant to Fresno State Digital Scholar the non-exclusive right to reproduce, translate (as defined in the next paragraph), and/or distribute your submission (including the abstract) worldwide in print and electronic format and in any medium, including but not limited to audio or video.

You agree that Fresno State may, without changing the content, translate the submission to any medium or format for the purpose of preservation.

You also agree that the submission is your original work, and that you have the right to grant the rights contained in this license. You also represent that your submission does not, to the best of your knowledge, infringe upon anyone's copyright.

If the submission reproduces material for which you do not hold copyright and that would not be considered fair use outside the copyright law, you represent that you have obtained the unrestricted permission of the copyright owner to grant Fresno State the rights required by this license, and that such third-party material is clearly identified and acknowledged within the text or content of the submission.

If the submission is based upon work that has been sponsored or supported by an agency or organization other than Fresno State, you represent that you have fulfilled any right of review or other obligations required by such contract or agreement.

Fresno State will clearly identify your name as the author or owner of the submission and will not make any alteration, other than as allowed by this license, to your submission. **By typing your name and date in the fields below, you indicate your agreement to the terms of this distribution license.**

Embargo options (fill box with an X).

Make my thesis or dissertation available to eCollections immediately upon submission.

Embargo my thesis or dissertation for a period of 2 years from date of graduation.

Embargo my thesis or dissertation for a period of 5 years from date of graduation.

Joshua P Kunnath

Type full name as it appears on submission

January 12, 2016

Date