CLEARvoz Journal

Center for Leadership Equity And Research
CLEAR

Examining the Role of Poverty in Teacher Grading Decisions

Joshua P. Kunnath, *Principal Research Center*
Mahmoud Suleiman, *California State University, Bakersfield*

Abstract
Poverty has a significant impact on the education of America's youth, causing an income achievement gap in American PreK-12 schools. A large amount of research has been done on poverty's effect on many aspects of schools, but few studies have addressed poverty's role on student grades. A century of grading research has shown that teacher grading practices are rarely an accurate representation of student academic achievement, but rather a construct of unclear meaning containing both objective and subjective factors. Because these practices are so difficult to understand, several studies have investigated teacher grading decision making to attempt a more analytic assessment of the process that produces these grades. The purpose of this study was to investigate teachers' grading decisions and the relationship to school poverty level in order to better understand the effectiveness of teacher grades in high-poverty schools. Using a causal-comparative design, the study was set in an urban California school district and used a sample of 251 high school teachers from 17 different high schools. A 35-item survey questionnaire was primarily used to determine the extent to which teachers used 17 different grading practices and were influenced by 13 different grading influences when creating report card grades. Results showed that teachers in low-poverty schools assigned significantly more A's than in mid- and high-poverty schools, while most grading practices and grading influences were consistent across school poverty levels. A discussion includes an interpretation of results within the context of grading literature, including the importance of four grading influences in high-poverty schools: *student success, teacher philosophy, school administrators,* and *student absenteeism.* The study concludes with recommendations for teachers and administrators in high-poverty schools to create and utilize effective report card grades in an effort to address the income achievement gap.

Keywords
grades, grading, poverty, achievement gap, decision making, secondary school teachers

Introduction

An income achievement gap existing between students of low and high income levels has been well documented in the United States (Hattie, 2009; Reardon, 2011, 2013; Sirin, 2005). Those documenting the gap have shown that students from families of higher income levels consistently display higher levels of academic achievement as measured by standardized tests as

compared to students from families of lower income levels. Support includes Hattie's (2009) synthesis of 499 studies on the topic, which found that student socioeconomic status (SES) had a moderate effect (d = .57) on academic achievement. But despite this relationship, student grade distributions across schools of all SES levels are remarkably similar (Randall & Engelhard, 2010; Zwick & Green, 2007).

Although some may interpret consistency in grade distributions across student SES levels as a sign of equitable practices, the lack of a relationship between grade distributions and SES can also be seen as problematic when one considers that teachers often intend for their grades to represent the same thing as standardized test scores: student academic achievement (Frary, Cross, & Weber, 1993; Kunnath, 2017). Student report card grades may be interpreted as a largely subjective construct when one considers that their creation often consists of an imprecise combination of nonachievement factors—in this study defined as noncognitive, subjective factors that include attitude, behavior, effort, and participation—and academic factors—defined as cognitive, objective factors primarily measured by individual test scores (Kunnath, 2017; McMillan, 2003; Randall & Engelhard, 2010; Reeves, 2011). Additionally, it is important to recognize that a likely reason for the homogeneity in grade distributions across schools is teachers' common use of nonachievement factors in creating report card grades (Guskey, 2015; Office of Educational Research and Improvement, 1994; Randall & Engelhard, 2009, 2010). But the question remains as to how teachers make decisions to create student grades. Further, educators must wonder how these decisions are affected by pressures that commonly exist in schools of high poverty, such as chronic student absenteeism, high student mobility, low standardized test scores, high teacher turnover, and less available learning time, among other factors (Ready, 2010; Rogers & Mirra, 2014; Rothstein, 2004; Stull, 2013).

This article examines these issues and explores the impact of poverty on teachers' grading practices. In particular, it presents the findings of the study conducted to investigate teachers' grading decisions and the relationship to school poverty level in an attempt to better understand the effectiveness of teacher grades—especially in high-poverty schools. The main research questions that underlie this investigation include: *(a) How does school poverty level affect teacher grading practices?* and *(b) How does school poverty level affect the influences on teacher grading?* Finally, implications are considered for teachers and administrators to achieve equitable grading processes while taking into account the income achievement gap.

Variation in Grading Practices

A century of grading research has shown student report card grades to be a highly variable measure of student achievement (Brookhart et al., 2016; Guskey, 2015; Schneider & Hutt, 2014). Although teachers commonly report using assessment results as a significant component of student grades, they also largely use nonachievement factors such as attitude, behavior, effort, and participation (Cross & Frary, 1999; Gusky, 2015; McMillan, Myran, & Workman, 2002; Reeves, 2008, 2011; Russell & Austin, 2010). Their methods for creating report card grades are often highly idiosyncratic, displaying high variation across schools, within schools, and even within individual classrooms (Duncan & Noonan, 2007; Guskey, 2002, 2009; McMillan, 2001).

As grading practices differ among school districts, schools, and teachers, course grades inevitably lose some of their meaning (Brookhart, 1994; Marzano, 2000). Variation in teacher grading practices were first documented more than 100 years ago (Starch & Elliott, 1912), and recent studies continue to report similar findings (Anderson, 2018; Brimi, 2011; Brookhart, 2013). 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. 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. Randall and Engelhard (2010) 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). In their study examining student sociocultural factors, grades, and SAT scores, Zwick and Green (2007) found fairly low variation of high school grade point average between schools (15.83%), while the variation between schools of SAT math (26.68%) and SAT verbal (26.06) scores was much higher. This seems to provide further evidence of grading variability, along with the use on nonachievement factors in grading. 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).

Teacher Grading Decision Making

One way to better understand teacher grading practices is to study the way in which teachers make their decisions about student grades. McMillan (2003) and McMillan and Nash (2000) created a grading decision-making model to explain the processes involved in teacher assessment and grading practices. The articles posited that these practices were the result of a rationale that was influenced by a combination of internal and external factors. This decision-making rationale refers to the logic and reasoning that teachers utilize when making grading decisions. The teacher grading decision-making model explains that three types of factors—internal factors (teacher knowledge, beliefs, expectations, and values), classroom realities (social promotion, absenteeism, disruptive behavior, and heterogeneity), and external factors (state accountability testing, district policies, and parents)—all contribute to the rationale that teachers employ to make grading and assessment decisions (McMillan, 2003).

While McMillan and Nash (2000) found that teachers often have a difficult time explaining this rationale, the model displays a number of possible factors at work, including two consistent findings from their studies: a *wide range of criteria* and *professional experience* (McMillan, 2003). The finding *a wide range of criteria* refers to teachers' belief that they should use multiple grading factors—both achievement and nonachievement—to fairly grade students (Brookhart, 1991; Cross & Frary, 1999; McMillan, 2003; Reeves, 2011). *Professional experience* refers to teachers' description of the development of their own grading practices through experience mostly their own personal experiences and informal learning from fellow teachers that occurred within the classroom and school site (Frisbie & Waltman, 1992; McMillan, 2003; Reeves, 2011).

Grading and Poverty

Teacher use of nonachievement factors when grading can distort the relationship between the report card grade and academic achievement. Thus far, a small amount of research has found a relationship between school poverty level and the factors (achievement and nonachievement) used to create student grades. Although the evidence is far from definitive, it appears that teachers of high-poverty schools are more likely to use greater amounts of nonachievement factors than those of lower poverty levels, as teacher grading practices seem to face greater influences from internal and external pressures in high-poverty schools (Agnew, 1985; Cauley & McMillan, 2000; Howley, Kusimo, & Parrott, 2000; Kelly, 2008; McMillan et al., 2002; Office of Educational Research and Improvement, 1994; Zwick & Himmelfarb, 2011). As a result, low achieving students may not be accurately identified, and unidentified students may lack opportunities to learn

deficient skills (McMillan, 2001). This is not only a matter of poor practice, but also a matter of equity (Office of Educational Research and Improvement, 1994).

While research on the impact of student poverty on grades has failed to illuminate the exact nature of the interaction, it seems that student grades in high-poverty schools are less valid measures of academic achievement than in schools of lower poverty levels (Agnew, 1985; Cauley & McMillan, 2000; Howley et al., 2000; Madon et al., 1998; Office of Educational Research and Improvement, 1994; Zwick & Himelfarb, 2011). Brennan et al. (2001) studied test equity by examining the relationship between teacher grades and high-stakes tests, analyzing 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.

Although several studies have analyzed teacher grading decision making (Cheng & Sun, 2015; Isnawati & Saukah, 2017; McMillan, 2003; McMillan and Nash, 2000, Kunnath, 2017), none to the knowledge of the authors has attempted to do so from an equity perspective. Thus, it is the goal of the authors to determine the extent to which school poverty level plays a role in teacher grading decisions in order to better understand the value of report card grades in high-poverty schools.

Method

Study Design

This study used a causal comparative design to investigate teacher grading decisions in schools of varied poverty levels (Best & Kahn, 2006). To do so, teacher grading decisions were compared by three school poverty groups: *low-poverty schools, mid-poverty schools,* and *high-poverty schools*. School poverty level was determined by the proportion of students eligible for the National School Lunch Program's free or reduced priced meals, often referred to as free or reduced-price lunch (FRPL), and hereafter in this article referred in this same way (see Domina et al., 2018). Low-poverty schools were defined as schools with 50% or less of students eligible for FRPL, 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.

Setting and Participants

The study was set in a large, ethnically-diverse urban California school district selected through convenience sampling. Survey participants were selected through comprehensive sampling, as all high school teachers of English, foreign language, mathematics, science, and social studies were invited to participate. A total of 251 teachers from 17 high schools completed the closed-response survey items, representing a 27% response rate.

Data Collection

Prior to instrument use, the study was approved by an internal review board. Cross-sectional data was collected with a self-administered Web-based questionnaire through SurveyMonkey. The survey, consisting of 35 items, was utilized to obtain quantitative data on several aspects of teacher grading decision making. The survey was organized into three components: teacher background, grading practices, and grading influences (Appendix A). Part 1 contained three items on teacher background, including teaching experience, predominant teaching assignment, and class level (college preparatory or not). This section was developed from the first group of a survey questionnaire items utilized by Cross and Frary (1999) and Frary et al. (1993) in their studies on teacher grading practices. Part 2 contained 19 items that were primarily designed to determine the methods that teachers used to conduct their grading practices. The first two items asked teachers of the percentage of A's and F's they assigned, and it this was adapted from Cross and Frary (1999) and Frary et al. (1993). The next 17 items asked teachers to use a 5-point Likert-type scale ranging from *not at all* to *completely* to answer the extent to which they used 17 common grading practices to create their final first semester grades. These items were modified from practices 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.

Part 3 consisted of 13 items that were used to determine the influences on teachers' grading practices. Survey items provided a five-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 of the 13 influences. The concept of *grading influences* was first developed by McMillan and Nash (2000) and refined by McMillan (2003) in their articles on teacher grading and assessment decision making, while the Likert-type scale came from the recommendations of Vagias (2006). The entire survey was piloted at a high school in a neighboring school district. Piloting prompted a number of survey revisions, including rewording of three items to enhance clarity and the deletion of an item on student social promotion that was perceived as irrelevant in the high school context.

Data Analysis

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 teacher self-reported use of 17 grading practices) by school poverty level. Goldwater and Nutt (1999) used the concept of an SGI in their study of the relationship between the compatibility of student and teacher backgrounds and teacher grading practices, but unlike the present study, subjectivity was measured by comparing student report card grades to final exam grades. The current study also used Kruskall-Wallis tests to compare the means of 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, Leech, Gloeckner, & Barrett, 2011). A multiple analysis of variance (MANOVA) test was used to compare the means of the 13 grading influences 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).

Results

Of the 915 teachers targeted in this study, 325 teachers (36%) began the survey, while 251 finished for a response rate of 27.4%. Survey responses varied by school poverty level group, as displayed in Table 1 below.

Table 1
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

Notably, despite the disproportionately high number of schools (n = 8) and teachers (n = 464) in the high-poverty group, the sample sizes of the three groups were 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 high-poverty schools group was because of the exceptionally low rate of one site in the group (6.7%) and the non-participation of another. All subsequent data will only reflect the data from the 17 participating sites.

Unlike the great range of survey completion by school site and poverty level, the types of teachers that responded to the survey were remarkably similar. Table 2 shows the teaching experience, major teaching assignment, and primary class level of teachers who responded to the survey by school poverty level.

Table 2
Demographics of Survey Participants by School Poverty Level

	School Poverty Level					
	Low		Mid		High	
Variable	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.

Teacher Grading Practices

Assigned A's and F's. The first aspect of teacher grading practices that was investigated was the proportion of A's and F's assigned by teachers for each school poverty level. Table 3 illustrates the mean, standard deviation, and median of teacher self-reported assigned A's and F's at each school poverty level. A self-reported score of 1 = *less than 3%*, 2 = *5 to 10%*, 3 = *11 to*

20%, and 4 = *more than 20%*. As shown in Table 3, 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 of self-reported assigned A's in mid-poverty schools (M = 2.58) were similar to the mean scores in high-poverty schools (M = 2.63), which falls somewhere between the *5 to 10%* and *11 to 20%* categories. Teachers of mid-poverty schools also displayed the highest mean for assigned F's (M = 2.01), at approximately *5 to 10%.*

Table 3
Means, Standard Deviations, and Medians Comparing Teacher Assigned Grades by School Poverty Level

	School Poverty Level								
	Low			Mid			High		
Grading Practice	*M*	*SD*	*Mdn*	*M*	*SD*	*Mdn*	*M*	*SD*	*Mdn*
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

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. To determine pairwise contrasts in the ANOVA results, 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 (p < .01, d = .53), according to Cohen (1988). Additionally, significant differences were found in teacher assigned 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 (p < .01, d = .52).

Use of 17 common grading practices. Next, teacher use of 17 common grading practices was analyzed by school poverty level. Descriptive results are shown in Table 4. A self-reported score of 1 = *not at all*, 2 = *slightly*, 3 = *somewhat*, 4 = *largely*, and 5 = *completely*. Across all school poverty levels (low, mid, and high), highest mean responses were found on *student academic achievement* (M = 3.84, 4.06, 4.00), *specific learning objectives mastered by students* (M = 3.78, 3.92, 3.85), and *student ability level* (M = 3.53, 3.63, 3.58). Lowest means were seen on *grade distributions of other teachers* (M = 1.20, 1.20, 1.08), *student performance compared to students from previous years* (M = 1.22, 1.31, 1.31), and *student disruptive behavior/conduct* (M = 1.23, 1.36, 1.40).

Across all school poverty levels (low, mid, and high), highest standard deviations were seen on *inclusion of zeros for incomplete assignments or assessments* (SD = 1.13, 1.29, 1.24) and *student participation and/or paying attention* (SD = 1.08, 1.09, 1.11). Lowest standard deviations were seen on *student extra credit for academic performance* (SD = .54, .66, .65) and *student extra credit for non-academic performance* (SD = .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* (SD = .35) and *student disruptive behavior/conduct in low-poverty schools* (SD = .55), which were lower than their respective comparison groups, and *student extra credit for non-academic performance in low-poverty schools* (SD = .59), which was higher than comparison groups.

Table 4
Means, Standard Deviations, and Medians Comparing Teacher Grading Practices by School Poverty Group

	School Poverty Level								
	Low			Mid			High		
Grading Practices	*M*	*SD*	*Mdn*	*M*	*SD*	*Mdn*	*M*	*SD*	*Mdn*
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
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

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. Mann-Whitney post hoc tests compared the three school poverty levels on *quality of student completed homework* and *student extra credit for academic performance* with a Bonferonni corrected p value of .017 to determine statistical significance (Morgan et al., 2011). For *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. For s*tudent 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 (Cohen, 1988).

A Subjective Grading Index (SGI) score was created for each survey respondent by averaging survey responses to the 17 grading practices. Two of the seventeen grading practices––*student academic achievement* and *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 produced an index in which a score of 1.0 represents minimum grading subjectivity 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, and the calculated score (α = .66) was found to be slightly below the minimum desired score (α = .70). The α score (α = .66) was based on standardized items because of the large variance in grading practice means and standard deviations (Morgan et al., 2011).

Table 5 displays the complete SGI results by school poverty level. The mean SGIs of all school poverty levels were similar, although low-poverty schools displayed a somewhat low standard deviation (SD = .26), and mid-poverty schools showed a somewhat low median (Mdn = 2.03). A one-way ANOVA was conducted to test for significant differences in SGI scores between school poverty levels; however, no statistical significant differences were found.

Table 5
Means, Standard Deviations, and Medians Comparing Subjective Grading Index by School Poverty Level

School Poverty Level	*M*	*SD*	*Mdn*
Low	2.10	.26	2.12
Mid	2.07	.39	2.03
High	2.14	.35	2.18

Influences on Teacher Grading

Thirteen influences on teacher grading were measured and analyzed by school poverty level with the use of thirteen survey items. As described in the methods section, a teacher self-reported score of 1.0 indicates that a factor was *not at all* influential on a teacher's grading practices, while a score of 5.0 indicates a factor was *extremely* influential. Table 6 illustrates descriptive results of the influences across all school poverty levels (low, mid, and high). The highest mean responses were seen on *philosophy of teaching and learning* (M = 4.00, 3.91, 3.88) and *desire to promote student understanding* (M = 3.93, 3.99, 3.94). Lowest means were seen on *parents* (M = 1.44, 1.44, 1.40) and *student disruptive behavior* (M = 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* (SD = 1.23, 1.24, 1.30), *student motivation and engagement* (SD = 1.15, 1.16, 1.13), and *student absenteeism* (SD = 1.11, 1.10, 1.15). Lowest standard deviations were seen in *parents* (SD = .67, .75, .70) and *student disruptive behavior* (SD = .78, .76, .73). Additionally, responses to the item *formal or informal school or district policies* displayed a notably lower

standard deviation (*SD* = .89) in low-poverty schools as compared to mid- (*SD* = 1.06) and high-poverty (*SD* = 1.08) schools.

Table 6
Means, Standard Deviations, and Medians Comparing Influences of Grading by School Poverty Group

	School Poverty Level								
	Low			Mid			High		
Influences of Grading	*M*	*SD*	*Mdn*	*M*	*SD*	*Mdn*	*M*	*SD*	*Mdn*
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

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 on 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 ($r = .613$), considered a medium to high effect size (Cohen, 1988), existed between the influences *formal or informal school or district policies* and *school administrators*. To address this possible source of multicollinearity, *formal or informal school or district policies* was eliminated as a dependent variable from the MANOVA test. The MANOVA tests found no statistically significant differences between the three school poverty levels. 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, *desire to accommodate student individual differences and needs* displayed a low *p*-value (.076) near the .05 significance level.

Discussion

Research Question 1

The first research question addressed the effect of school poverty level on teacher grading practices. Overall, the findings demonstrated low-level effects.

Grade distributions. Results showed some grade distribution variation across school poverty levels, as teachers of low-poverty schools assigned the greatest proportion of A's and among the lowest proportion of F's. Specifically, teachers of low-poverty schools were found to assign 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. A possible explanation for more assigned A's and fewer F's in low-poverty schools is the pressure that teachers receive from parents, administrators, and students in these schools to assign higher grades. Because most students in low-poverty schools come from homes of higher SES levels, it is more likely that they have parents who are actively involved in their education, leading to a greater likelihood of increased teacher contact and subsequent pressure on teacher grading practices. Conversely, students of high-poverty schools often come from low-SES households and are less likely to have parents who are actively involved in school, lowering the likelihood of adding pressure on teachers to alter grading practices (Horvat, Weininger, & Lareau, 2003; Lee & Bowen, 2006). These results differ somewhat from previous studies, which show that student grade distributions are fairly uniform across school poverty levels (Office of Educational Research and Improvement, 1994; Randall & Engelhard, 2010; Zwick & Green, 2007). However, a report from the Office of Educational 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, Borg, & Stranahan, 2012; Dahl & Lochner, 2012; Duncan & Murnane, 2014; Heckman, 2006; Reardon, 2011; Sirin, 2005; Stull, 2013), the proportion of failing students in low-poverty and high-poverty schools appears to be fairly similar. This may be attributed to the great attention and significance placed on teachers assigning F's. Wiley (2011) found that teachers' varying use of nonachievement factors often depended upon student ability and achievement level, as teachers were found to use greater proportions of effort in grading for low ability or low-achieving students. Additionally, several studies, such as Bonner and Chen (2008); Cizek, Fitzgerald, and Rachor (1995); and Guskey (2002), found teachers to use whatever combination of achievement and nonachievement factors that resulted in the highest student grades. Further, in cases of borderline grades, teachers may be lenient in rounding up F's to passing grades because of the stigma put on both the student and the teacher when a student fails a class. This pressure can serve 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 (Randall & Engelhard, 2010; Sun & Cheng, 2013; Tierney, Simon, & Charland, 2011).

Objective and subjective practices. Teachers in all school poverty levels seemed to use a combination of objective, or achievement-based, and subjective, or nonachievement-based, practices. Of the 17 different grading practices included in the survey, teachers across all school poverty levels collectively reported they *largely* used the only two practices that 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). However, teachers in all school poverty levels also displayed a significant use of subjective grading

practices. Subjective practices most used were *student ability level*, *student 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. High survey scores for objective practices seems to indicate that teachers recognize the importance of using achievement factors in their grades; however, mid-level scores for subjective practices implies that teachers also perceive nonachievement factors to be an important part of student grades. These results were similar to recent studies reviewed by Brookhart et al. (2016), who found that teachers' grades commonly include both cognitive and noncognitive factors.

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 these differences across school poverty levels were inconsistent, it is difficult to make any conclusions about poverty's role on these practices. Despite this inconclusiveness, the existence of some impact of school poverty level on teacher subjective grading practices is apparent and deserves further study.

SGI scores constructed from the survey, which quantified the subjectivity of teacher grading practices, indicated that teachers across all poverty levels uniformly scored between 2.07 and 2.14 on the 5-point scale. However, considering the fact that grading experts recommend *no* use of subjective practices (Guskey, 2015), equivalent to an SGI score of 1.0, these results may be interpreted as contributing a relatively high level of subjectivity to grade meaning. In other words, once SGI scores rise above a level of 1, grade meaning shifts from one of student academic achievement to one of an indeterminate mixture of academic achievement and various nonachievement factors. The higher the score, the higher the subjectivity and the less clarity in grade meaning.

Teachers' reported use of the 15 subjective grading practices were consistent across poverty levels. The use of these types of practices is similar to those first described by Brookhart (1991) and later by others (e.g., Cizek et al., 1995; 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). In Kunnath's (2017) study of teacher grading decisions, he found that teachers often relied more heavily on subjective non-achievement factors when they felt heavy weighting of test grades would significantly lower overall report card grades. Teachers explained that they used non-achievement factors more often in lower-level classes (i.e., non-college preparatory), while they relied more heavily on achievement factors in higher-level classes (e.g., Honors, GATE, AP). It stands to reason that in schools in which fewer students are enrolled in higher-level classes, grades may be less accurate than in schools in which a greater number of students are enrolled in these higher-level classes. Because high-poverty schools often provide students with less access to these higher-level classes (Bittman, Davies, Russell, & Goussakova, 2017; Kolluri, 2018), teachers in these schools may be more likely to use hodgepodge grading practices.

Research Question 2

The second research question addressed the effect of school poverty level on influences on 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. However, four influences were interpreted as important factors across all school poverty levels: *student success, teacher philosophy, school administrators,* and *student absenteeism.*

Student success and teacher philosophy. The influences *desire for student success* and *philosophy of teaching and learning* were interpreted as significant influences on teacher grading, with teacher survey scores of 3.0 (somewhat influential) or higher across all school poverty levels. *Desire for student success* registered a mean influential score of 3.31 (between somewhat and very influential) across all poverty groups. McMillan (2003) and McMillan and Nash (2000) found similar results and termed the phenomenon "pulling for students," explaining that 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). Other studies have reported similar findings, explaining that teachers often use grading practices that result in the highest possible grades for students (Bonner & Chen, 2008; Cizek, et al., 1995; Guskey, 2002). The influence *philosophy of teaching and learning* was scored as the highest of all 13 influences on grading, with a mean score of 3.93 (just below very influential). This seems to indicate that although there are many influences on teacher grading practices, teachers feel their grades are purposefully constructed to align to their own philosophy. Interestingly, these two high-scoring influences likely produce converse effects on grades. While *philosophy of teaching and learning* is more likely to result in objective grading practices, as teachers often value the use of summative assessments when creating report card grades (Frary et al., 1993; Kunnath, 2017), *desire for student success* most likely results in subjective practices for reasons mentioned above. The high scores of these two influences seem to indicate that the relative degree of grading objectivity largely depends upon which influence is stronger for the teacher at the time of the grading decision. In high-poverty schools, the greater occurrence of high-need students may result in teachers more affected by *desire for student success* than *philosophy of teaching and learning*.

School administrators and student absenteeism. Two additional influences—*school administrators* and *student absenteeism*—were 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*. Teachers of high-poverty schools scored this influence highest (1.90) compared to mid-poverty schools (1.76) and low-poverty schools (1.70). In the case of this influence, any score above 1.0 (not at all influential) may be interpreted as a significant because of the power potential of school administrators on grading decision making. This interpretation aligns with the findings of a few studies of pressures that administrators often place on teacher grading practices to limit failing grades and produce report card grades that fall within a normal distribution (Agnew, 1985; Cross & Frary, 1999; Iacus & Poro, 2011). Because of this phenomenon, teachers with greater numbers of students who are less inclined for academic success are more likely to experience pressure from administrators. As high-poverty schools often 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*. Although *student absenteeism* was not significantly different across groups, it was deemed a significant influence because of its likelihood of disproportionately affecting schools across poverty levels. Because student absenteeism is a larger issue in high-poverty schools (Rogers & Mirra, 2014; Rothstein, 2004), 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. Rogers and Mirra (2014) and 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. The additive effect of these instances in high-poverty schools likely adds to the subjectivity and inaccuracy of student grades.

Notwithstanding some of the study limitations, data trends and findings have shed significant light on the place of poverty and its impact on teacher grading. This uncharted territory is worth exploring with further research and study. The research results in this area will inevitably have direct implications for narrowing the income achievement gap and help teachers and administrators provide better ways of equitably accounting for what students know and are able to do.

Conclusion and Recommendations

Educators within high-poverty schools should openly discuss the invisible pressures that can affect teacher grading practices to maximize the clarity of grade meaning. By engaging in school-wide discussions with faculty and administrators, schools can collectively make effective grading decisions entirely from their philosophy of teaching and learning—an ideal influence. When schools can create grades that truly represent student learning, all subsequent conversations about student grades also become conversations about student learning—something very different from what often occurs. Such a proactive approach can help to minimize reactive pressures coming from administrators that arise after teachers create and submit report card grades, likely occurring because administrators are unaware of a teacher's specific grading decision making process. Further, by using only measures of academic achievement to create student report card grades, teachers of these schools can eliminate subjective influences such as *desire for student success,* which obscure grade meaning. However, to address the needs and concerns that teachers often feel when "pulling for their students" in their grading, schools using these objective practices must provide extensive school-wide remediation structures. These structures should provide learning support for students who have low grades, likely D's or F's, which implies that they failed to achieve an adequate level of learning of learning targets. Crucially, these supports should provide additional learning opportunities without penalizing students for taking longer to achieve proficiency. This includes ample opportunities for absent students to learn skills and concepts and display their learning to prevent the need for the use of the zero grade. A grade of zero, after all, implies zero learning, but this meaning is rarely accurate.

Although the suggestions above are strong practices for any school, they are especially important in high-poverty schools. In these schools, report card grades that accurately represent student learning can help educators to better determine student proficiency and deficiency on essential learning targets, which can help to lead remediation efforts. Once schools can accurately determine their remediation needs, they can better create structures and allocate resources that target these needs—work that is crucial in addressing the income achievement gap.

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.

Anderson, L. W. (2018). A critique of grading: Policies, practices, and technical matters. *Education Policy Analysis Archives, 26*(49), 1-31.

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)

Best, J. W., & Kahn, J. V. (2006). *Research in education* (10th ed.). Boston, MA: Pearson.

Bittman, B., Davies, A., Russell, W., & Goussakova, E. (2017). Advance Placement and the achievement gap in the 21st century: A multiple linear regression of marginalized populations in AP enrollment. *Research in Social Sciences and Technology*, *2*(2), 36-58.

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

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.

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.

Brimi, H. M. (2011). Reliability of grading high school work in English. *Practical Assessment, Research & Evaluation, 16*(17), 1-12.

Brookhart, S.M. (1991). Grading practices and validity. *Educational Measurement: Issues and Practice, 10*(1), 35-36.

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

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

Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., … Welsh, M. E. (2016). A Century of Grading Research: Meaning and Value in the Most Common Educational Measure. *Review of Educational Research*, *86*(4), 803–848.

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.

Cheng, L., & Sun, Y. (2015). Teachers' grading decision making: Multiple influencing factors and methods. *Language Assessment Quarterly, 12*, 213-233.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Cizek, G. J., Fitzgerald, S. M., & Rachor, R. E. (1995). Teachers' assessment practices: Preparation, isolation, and the kitchen sink. *Educational Assessment, 3*(2), 159-179.

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

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.

Domina, T., Pharris-Ciurej, N., Penner, A. M., Penner, E. K., Brummet, Q., Porter, S. R., & Sanabria, T. (2018). Is free and reduced-price lunch a valid measure of educational disadvantage? *Educational Researcher, 47*(9), 539-555.

Duncan, G. J., & Murnane, R. J. (2014). Growing income inequality threatens American education. *Phi Delta Kappan*, *95*(6), 8-14.

Duncan, R. C., & Noonan, B. (2007). Factors affecting teachers' grading and assessment practices. *Alberta Journal of Educational Research, 53,* 1–21.

Frary, 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.

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.

Goldwater, O. D., & Nutt, R. L. (1999). Teachers' and students' work-culture variables associated with positive school outcome. *Adolescence*, *34*(136), 653-664.

Guskey, T. R. (2002, April). *Perspectives on grading and reporting: Differences among teachers, students, and parents*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Guskey, T. R. (2009, April). *Bound by tradition: Teachers' views of crucial grading and reporting issues*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Guskey, T. R. (2015). *On your mark: Challenging the conventions of grading and reporting*. Bloomington, IN: Solution Tree Press.

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.

Horvat, E., Weininger, E., & Lareau, A. (2003). From social ties to social capital: Class differences in the relations between schools and parent networks. *American Educational Research Journal, 40*(2), 319-351.

Howley, A., Kusimo, P., & Parrott, L. (2000). Grading and the ethos of effort. *Learning Environments Research*, *3*(3), 229-246.

Iacus, S. M., & Porro, G. (2011). Teachers' evaluations and students' achievement: a 'deviation from the reference' analysis. *Education Economics*, *19*(2), 139-159.

Isnawati, I., & Saukah, A. (2017). Teachers' grading decision making. *TEFLIN Journal, 28*(2), 155-169.

Kelly, S. (2008). What types of students' effort are rewarded with high marks? *Sociology of Education*, *81*(1), 32-52.

Kolluri, S. (2018). Advanced Placement: The dual challenge of equal access and effectiveness. *Review of Educational Research*, *88*(5), 671–711.

Kunnath, J. P. (2017). Teacher grading decisions: Influences, rationale, and practices. *American Secondary Education, 45*, 68-88.

Lee, J., & Bowen, N.K. (2006). Parent involvement, cultural capital, and the achievement gap among elementary school children. *American Educational Research Journal, 43*(2), 193-218.

Leech, N. L., Barrett, K. C., & Morgan, G. A. (2011). *IBM SPSS for intermediate statistics: Use and interpretation* (4th ed.). New York, NY: Routledge.

Marzano, R. J. (2000). *Transforming classroom grading*. Alexandria, VA: ASCD.

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. *Journal of Educational Research, 95*, 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.

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.

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.

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

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.

Reardon, S. F. (2013). The Widening Income Achievement Gap. *Educational Leadership, 70*(8), 10-16.

Reeves, D. B. (2008). Leading to change/effective grading practices. *Educational Leadership, 65*(5), 85-87.

Reeves, D. B. (2011). *Elements of grading: A guide to effective practice*. Bloomington, IN: Solution Tree Press.

Rogers, J., & Mirra, N. (2014). *It's about time: Learning time and educational opportunity in California high schools*. Los Angeles, CA: UCLA IDEA.

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.

Russell, J. A., & Austin, J. R. (2010). Assessment practices of secondary music teachers. *Journal of Research in Music Education, 58*, 37–54.

Schneider, J., & Hutt, E. (2014). Making the grade: A history of the A – F marking scheme. *Journal of Curriculum Studies, 46*(2), 201-224.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*(3), 417-453.

Starch, D., & Elliott, E. C. (1912). Reliability of grading high-school work in English. *The School Review, 20*(7), 442-457.

Stull, J. C. (2013). Family socioeconomic status, parent expectations, and a child's achievement. *Research in Education*, *90*, 53-67.

Sun, Y., & Cheng, L. (2013). Teachers' grading practices: Meaning and values assigned. *Assessment in Education, 21*, 326–343.

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

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.

Wiley, C. R. (2011). *Profiles of teacher grading practices: Integrating teacher beliefs, course criteria, and student characteristics* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 887719048)

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.

Appendix A

Teacher Grading Practices and Influences Questionnaire

Part 1: Teacher Background

1. How many total years have you been teaching?
   Less than 3 3 to 6 7 to 10 More than 10

2. What was your major teaching assignment (class taught with the most sections) during the second semester of last school year?
   English Foreign Language Mathematics Science Social Studies

3. What proportion of your classes were college preparatory (i.e., CP, GATE, Honors, or AP)?
   Half or more Fewer than half

Part 2: Grading Practices

4. 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% 5-10% 11-20% More than 20%

5. 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% 5-10% 11-20% More than 20%

6. 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

7. To what extent were your final second 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

8. 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

9. 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

10. 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

11. 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

12. 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

13. 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 classes?
    1. Not at all (2) Slightly (3) Somewhat (4) Largely (5) Completely

14. To what extent were your final second semester student grades in your major teaching assignment based on student performance compared to students from previous years?
    1. Not at all (2) Slightly (3) Somewhat (4) Largely (5) Completely

15. 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

16. 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, and Fs?
    1. Not at all (2) Slightly (3) Somewhat (4) Largely (5) Completely

17. 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-text indicators for borderline grades?
    1. Not at all (2) Slightly (3) Somewhat (4) Largely (5) Completely

18. 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

19. To what extent were your final second semester 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

20. 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

21. 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

22. 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

Part 3: Grading Influences

23. 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

24. 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

25. 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

26. 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

27. How influential is student motivation and engagement on your grading practices?
    1. Not at all (2) Slightly (3) Somewhat (4) Very (5) Extremely

28. How influential is state standardized testing on your grading practices?
    1. Not at all (2) Slightly (3) Somewhat (4) Very (5) Extremely

29. 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

30. How influential are your administrators on your grading practices?
    1. Not at all (2) Slightly (3) Somewhat (4) Very (5) Extremely

31. How influential are parents on your grading practices?
    1. Not at all (2) Slightly (3) Somewhat (4) Very (5) Extremely

32. How influential is student absenteeism on your grading practices?
    1. Not at all (2) Slightly (3) Somewhat (4) Very (5) Extremely

33. How influential is student disruptive behavior on your grading practices?
    1. Not at all (2) Slightly (3) Somewhat (4) Very (5) Extremely

34. How influential is differing student ability level on your grading practices?
    1. Not at all (2) Slightly (3) Somewhat (4) Very (5) Extremely

35. 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