RACIAL DIFFERENCES IN THE PSYCHOMETRIC PROPERTIES OF GRADES: ARE THE GRADES OF NON-WHITE STUDENTS MORE VARIABLE?

A Thesis

by

CHRISTEN LORRAINE DOVALINA

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Christopher M. Berry
Committee Members,	Winfred Arthur, Jr.
	Oi-Man Kwok
Head of Department,	Douglas W. Woods

May 2014

Major Subject: Psychology

ABSTRACT

Although college grades are used in the assessment of academic performance and in the determination of the criterion-related validity of tests designed to predict academic performance, there exists virtually no systematic empirical research investigating race differences in the psychometric properties of college grades. One reason to suspect that the psychometric properties of grades might display subgroup differences is that current theories of racial bias suggest some professors might grade non-White students more harshly, but others might grade them more leniently, meaning that grades for these students might be more variable and less intercorrelated throughout the college career than the grades of White students. In addition to the possibility of racial bias, there are a number of other race-related factors which could also increase grade variability for non-White students.

In the present study, I use a large educational data set, which includes the first 60 grades received by more than 150,000 students from 41 different colleges, to determine (1) whether White students' grades are more intercorrelated than non-White students' grades, (2) whether the average within-person standard deviation of grades is higher for non-White students than for White students, (3) whether the patterns of race differences in within-persons grade variability are the same for both public and private schools, (4) whether the increased variability of non-White student grades can be accounted for by other alternative explanatory factors (SAT scores, socioeconomic status, high school grade point average, course load, English as best language, and college GPA), and (5)

whether race differences in within-persons grade variability is a persistent trend across all four years of college. I found that Asian, Black, and Hispanic students did tend to have more variable grades than White students; however, this greater variability was almost completely accounted for by race differences in course load, high school GPA, English as best language, SES, SAT scores, and cumulative college GPA. Although the present study does not provide definitive evidence for or against the possibility of bias in grading, it suggests that, if there is bias in grading, it is not expressing itself by increasing the variability of non-White students' grades.

DEDICATION

For my three amazing brothers, Justen, Jason, and Jackson Kontny. You inspire me more than you know.

ACKNOWLEDGMENTS

I am grateful to my advisor and thesis committee chair, Dr. Christopher M. Berry, for his constant support and encouragement, seemingly endless patience, insightful guidance, and diligent attention to detail throughout this process. I would also like to thank my committee members, Dr. Winfred Arthur, Jr. and Dr. Oi-Man Kwok, for their thoughtful and helpful feedback on this project. Thank you.

Thank you also to my mother, Mary Beth Justen, for being such an excellent role model for me my whole life. Her own strong work ethic and high standards for excellence are constant sources of motivation for me, and her love, support, and prayers have seen me through every milestone. Thanks, Mom.

And of course, thank you, thank you, thank you to Mario Dovalina. I cannot begin to list all the ways that he has contributed not only to this thesis and my academic progress, but also to my personal growth and general well-being. I couldn't have asked for more in a partner and friend.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

Because college grades are often used as performance criteria when assessing academic performance or the criterion-related validity of tests designed to predict academic performance (e.g., Culpepper & Davenport, 2009; Pearson, 1993; Sackett et al., 2009; Vartanian et al., 2007; Yazedjian, Toews, & Navarro, 2009; Young, 1993; Zwick & Sklar, 2005), as well as to predict future academic or job performance (Vaquera & Maestas, 2008), it is important that college GPA be both reliable and valid; however, there is little existing empirical research on the psychometric properties of student grades, or whether these properties display race-based subgroup differences. Typically, college students are given individual grades for each class they take, and a grade point average, or GPA, is calculated on a scale of 0.0 - 4.0 for each student. The individual grades used to calculate overall GPA can be seen as items in the GPA measure. The present study focuses on two psychometric properties of grades: (1) the within-persons variability of grades across the college career (i.e., how variable is a given student's grades throughout college?) and (2) the within-persons intercorrelation of grades across the college career (i.e., how predictive is one college grade of other college grades for students?). Of particular interest in the present study is whether, and why, these two psychometric properties might differ for White and non-White students.

Understanding whether there are racial/ethnic subgroup differences in the psychometric properties of college grades is important for at least two reasons. First, if the

grades of non-White students were more variable and/or less intercorrelated across the college career, then it is possible that this could be due to some form of bias or discrimination in grading practices. The identification of such differences, or the lack thereof, in psychometric properties of grades would be an important first step toward understanding whether discriminatory grading practices are an issue needing further investigation. Second, there is a large literature that uses college grades as the performance criterion when determining whether college admissions tests (e.g., the SAT) exhibit predictive bias (i.e., systematically under- or over-predict the college grades of any racial/ethnic subgroup; Young & Kobrin, 2001). Such predictive bias analyses entail the assumption that the performance criterion (i.e., college grades) is itself unbiased, a topic that has received surprisingly little empirical research. The present study addresses this overlooked topic.

This all raises the question of why the grades of non-White students might be more variable and less intercorrelated across the college career than the grades of White students. There are a number of possibilities, of which racial bias is one. For the purposes of this thesis, I define racial bias in grading as students being graded differently based on their race, such that race accounts for variance in grades that is independent of true academic performance. According to current theories of racism, racial bias is one source of variance that might affect the grades of non-White students in different ways, depending on factors such as the beliefs held by their professors and the amount of ambiguity in grading decision rules (e.g., Gaertner & Dovidio, 2005; Katz, Glass, & Cohen, 1973; Katz, Wackenhut, & Hass, 1986; Pettigrew, 1979). This means that if

instructors are aware of the race of their students, some instructors might grade non-White students more harshly, but others might grade them more leniently, than White students, which could cause grades for these students to be more variable across the college career, or cause individual grades to be less predictive of one another. For example, if non-White students' grades are affected by racial bias, the average non-White student might have a wider range of grades across the college career (e.g., a White "B" student might get grades of 3.0, 3.0, and 3.0 across three courses, but a non-White "B" student might get grades of 2.7, 3.0, and 3.3.) and/or the average non-White students' grade in one college course might be less predictive of grades they earn in other courses, compared to the average White student.

Of course, racial bias in grading is not the only factor that could cause race differences in the variability and intercorrelation of college grades. For instance, racial differences in grade variability might be accounted for by other relevant factors that are related to (i.e., differ by) race, such as college preparation or socioeconomic status (SES). If factors such as race differences in college preparation or SES can account for race differences in grade variability, then this would be evidence unsupportive of racial bias in grading causing grades of non-White students to be more variable. However, if these factors cannot account for race differences in grade variability, the possibility of racial bias in grading remains.

The present study addresses these issues. First, the theoretical and empirical evidence supportive of racial bias in grading is reviewed, focusing on how such bias would cause the grades of non-White students to be more variable, and less

intercorrelated, across the college career. Then, factors other than bias in grading (e.g., college preparation, SES) that could cause race differences in grade variability/intercorrelations are reviewed. The present study then provides an empirical test of these propositions. I first calculated the intercorrelations of students' grades across the college career for each of four racial subgroups (Asian, Black, Hispanic, and White), in order to determine whether there were differences between races in the intercorrelations of grades. I then calculated the average within-persons standard deviation of grades for each subgroup, in order to determine if there were differences between races in grade variability. I then used a regression model to control for known race-related factors such as those mentioned above in order to determine if differences between races in variability, or the lack thereof, remained.

Theoretical Explanations for Racial Bias in Grading

Although differential variability in non-White compared to White grades could be due to a number of different factors, psychological theories of racial bias suggest a number of different mechanisms through which prejudices might impact subjective grading decisions. Theories of racial bias that would support differential variability of grades based on race include Aversive Racism Theory, Ambivalent Racism Theory, and Ultimate Attribution Error. Although an assumption of constant racial bias affecting all non-White students would only affect the mean GPAs of these students, effectively making them either higher or lower than those of White students, current theories suggest that the effects of bias change depending on the biased individual, the victim of bias, and the situational factors surrounding their interaction. This is the justification behind the idea

that bias would make the grades of non-White students more variable than those of White students. For non-White students, variability in professor harshness or leniency in subjective grading decisions would in turn cause grades across different courses to be more variable. Because these factors would have no effect on White students, bias could serve to make non-White students' grades more variable than those of White students.

This all entails that there is subjectivity in grading and that instructors are aware of the race of their students. Of course, there are instances where at least one of these conditions will not be met (e.g., a course with all multiple-choice exams that do not lend themselves to subjectivity in grading, a course with hundreds of students where the instructor cannot keep track of individual students or their races). However, there are certainly many college courses that include assignments that allow subjectivity in grading (e.g., writing assignments, essay exams) and that have few enough students that instructors could be aware of the race of some of their students. The fact that college courses differ in these characteristics further supports the idea that grades of non-White students might be more variable than those of White students (i.e., non-White students will encounter courses with and without these characteristics).

Ambivalent Racism Theory. Ambivalent Racism Theory (Katz, Glass, & Cohen, 1973; Katz, Wackenhut, & Hass, 1986) suggests that many White individuals simultaneously hold both positive and negative opinions about non-Whites. This may be due in part to the American value system, which emphasizes both egalitarianism and individualism. The egalitarian value promotes democracy and humanitarianism, but the value of individualism promotes the importance of personal freedom, self-reliance, and

hard work. In many cases, these values are in opposition to one another, and can foster racial ambivalence, which is the coexistence of two opposing opinions of non-White individuals. For example, from an egalitarian perspective, a White individual would view the non-White individual, especially those who have been persecuted, discriminated against, or are seen as coming from underprivileged backgrounds, favorably, and wish to help them. On the other hand, the value of individualism might make White individuals view non-White individuals more harshly, thinking that any hardship they are in could be alleviated by hard work and self-reliance. These two conflicting perspectives can cause individuals to focus on different characteristics of people they interact with, and might come into play in response to specific cues. Specifically, ambivalent racism theory suggests that racial ambivalence will create a conflict within individuals, such that responses to the subject of the conflicting views will be amplified. Therefore, responses to positive behaviors or other cues will be more positive, and responses to negative behaviors and cues will be more negative.

Ambivalence toward non-Whites, and the resulting amplification of either positive or negative attitudes and behaviors toward them, could impact the grades of non-White students. Student performance is naturally variable to some extent, and racial ambivalence of professors could act to amplify the impact of this variability on resulting grades. For non-White students, good work would receive an especially high grade, and poor work would receive an especially low grade, compared to comparable work from White students. Thus, given the same variability in "true academic performance", non-White students would have more variable grades across assignments and courses. Additionally,

the grades of non-White students would be less intercorrelated (i.e., the grade a student gets in one course has less bearing on grades they receive in other courses) because grades for any given course may or may not reflect construct-irrelevant variance arising from ambivalent racism.

Ultimate Attribution Error. Ultimate Attribution Error (Pettigrew, 1979) proposes that individuals who hold prejudices toward members of a group will attribute their behaviors to different causes than they would for individuals of their own ingroups. Specifically, negative actions will be more likely to be seen as due to dispositional causes, but positive actions will be more likely to be seen as due to being an exceptional case, luck, high motivation or effort, or situational factors. For example, a student might fail to turn in an assignment for any number of reasons. It might be due to anything from a lack of conscientiousness to an emergency injury that made completing the assignment physically impossible. Ultimate Attribution Theory suggests that, if a White professor held a prejudice against non-White students, then if a White student were to fail to turn in an assignment, the professor would more likely assume that something extraneous must have happened to prevent it. However, if a non-White student were to fail to turn in the same assignment to the same professor, then that professor would more likely assume that student is lazy, or does not care about the class.

This could lead to non-White students' grades being more variable if professors were to make assumptions early on about performance and grade later assignments more harshly, or if an important grade were not granted leniency over an extenuating circumstance. This would mean that those unsystematic situational changes that affect all

students would have a greater effect on non-White students, because a negative behavior would be most likely judged the most harshly, and positive behaviors might fail to build that student as much trust as they would for a White student. Also, not only would non-White student grades be more affected by variability in their own circumstances, but this effect would be different from class to class. So, for example, perhaps a consistently underperforming student would be judged extra harshly from the start by some professors, but would still be given the benefit of the doubt by those who do not hold such biases. Therefore, grades could vary based on an increased effect of life circumstances, or based on whether or not the professor for a given course held prejudices.

Aversive Racism Theory. Another recent theory that would expect racial bias to contaminate grades is Aversive Racism Theory (e.g., Gaertner & Dovidio, 2005), which proposes that racism is prevalent today, but in less straightforward, obvious ways than in the past. Aversive racists do not regard themselves as prejudiced; outwardly, they sympathize with non-Whites as victims of injustice and support the principle of racial equality. However, aversive racists also possess unconscious negative beliefs about non-Whites, along with feelings of uneasiness over potentially being perceived as racist. Discrimination will therefore tend to occur in subtle, indirect, or rationizable ways, such as when decision guidelines are weak or vague (Gaertner & Dovidio, 2005). For example, when a White and non-White student both clearly perform very well on an assignment (thus, the grading decision is not vague), an aversive racist professor would be expected to grade the two students similarly. However, an aversive racist professor might grade a non-White students' mediocre work (for which the grading decision is vaguer) more

harshly than the same work from a White student, rationalizing the especially harsh grading because the paper was not particularly good in the first place. The aversive racism mechanisms could contaminate grades of non-White students with racial bias in addition to true performance variance. The grades of non-White students would therefore be more variable than the grades of White students, because a portion of their grades would be determined by whether or not their professor did hold negative beliefs about non-White students. This also suggests that in situations where grading is more subjective, grades would be more easily influenced by professor prejudices, whereas situations where grading is more objective would lessen the likelihood that prejudice would play a role in grading.

Conclusions. Theories of racial bias suggest that White individuals perceive and interact with non-White individuals differently than they would individuals from their own racial subgroup. The specific mechanism through which bias might affect grades varies across different theories. Professors might grade good and bad work as more extremely good or bad than it actually is due to simultaneously holding two conflicting beliefs about non-Whites. They might be more likely to attribute the quality of a non-White student's work to different causes than they would a White student's work, causing them to grade either more leniently or more harshly, depending on the situation. They might also discriminate against non-White students on assignments requiring some subjective judgment, without even being aware that they are doing it. Although the specific mechanisms through which racial bias might affect grades are uncertain, the effects of any of these mechanisms would lead to more variable grades for non-White

students than for White students. They would also lead to the grades of non-White students being less intercorrelated than the grades of White students. If racial bias does exist among college professors, then the differences in biases among different professors for different classes would mean that, for non-White students, grades would be based at least in part on whether the professor for a given class was biased, and in which direction that bias expressed itself. These differences would mean that similar performance in two different classes would be more likely to result in two different grades for a non-White student than for White students who would not be affected by racial biases.

Relevant Evidence in Educational Contexts

Despite the potential importance of this issue and the theories of racism that would posit that racial bias affects college grades, there has been a surprising dearth of research on racial bias in grading. However, this is not to say the idea of racial bias has been ignored in educational contexts. Qualitative accounts of racial discrimination in educational settings (e.g., Feagin, Vera, & Imani, 1996; Pinel, Warner, & Chua, 2005; Feagin, 1992) testify to the need for systematic research regarding racial bias in this area, in order to shed light on the amount and form of racial bias impacting the grading process. Additionally, a voluminous literature exists on differences in (1) the expectations teachers have regarding White versus non-White students and (2) the nonverbal behaviors teachers exhibit toward White versus non-White students. Although differences in the expectations and non-verbal behaviors teachers express toward non-White students are not direct evidence of racial bias affecting students' grades themselves, they are compatible with the existence of bias in educational settings.

Qualitative Accounts of Discrimination in Education. A number of qualitative studies have demonstrated that non-White students do feel they have experienced discrimination due to their stigmatized status in predominately White institutions. Although these reports of experiences of discrimination are not evidence that discrimination actually took place, they are evidence that the college experience does vary for students based on race. For example, Feagin (1992) conducted a qualitative field research study in which Black students, administrators, and faculty members were interviewed in order to determine whether racial discrimination may still be more prominent in predominately White universities than is currently believed. In this study, 28 students from predominately White universities across 14 different cities were interviewed regarding their college experiences, and in particular experiences of racial discrimination. Students reported experiencing discrimination from a number of different sources, including faculty, staff, and other students. Some of the identified discriminatory practices against Black students included aggression, exclusion, dismissal of subculture, and typecasting (Feagin, 1992).

Pinel, Warner, and Chua (2005) measured students' levels of stigma consciousness before and after they began attending a predominately White university. Stigma consciousness is the extent to which targets focus on their stereotyped status and believe it pervades their life experiences. Stigma consciousness is both dispositional and situationally induced; individuals who experience racial discrimination, or are otherwise made more aware of their race, can be expected to have an increase in stigma consciousness. This study divided participants into stigmatized (Black $n = 20$; Hispanic n

= 44) and non-stigmatized (White *n* =113; Asian *n* = 32) groups, and gave them a survey that asked about their experiences of stigma consciousness before they began attending the university and since they began attending the university. Individuals in stigmatized groups reported increased stigma consciousness in the time after arriving at the university compared to before, but non-stigmatized individuals did not.

These types of studies cannot be seen as proof that there is bias in grading. In fact, they cannot be seen as proof that there is bias at all; rather, they focus on individuals' experiences and/or perceptions of bias. Nevertheless, the fact that individuals from stigmatized groups do report feeling the effects of discrimination, and seem to become more conscious of their membership in stigmatized groups after attending a predominately White university, suggests this is an issue that is important for future research. Although the mechanisms at play may be unclear, these studies suggest that the college experience may be different for non-White students than it is for White students. Although it is unclear why these differences exist and why exactly students feel discriminated against, there is a possibility that there are biases held by professors and other students.

Teacher Expectations and Nonverbal Behaviors Toward Non-White Students. Research on teacher expectations and nonverbal behaviors has come to conflicting conclusions; some find more positive attitudes and behaviors toward White students, but others find no race differences, or even more positive attitudes and behaviors toward non-Whites. However, the majority of existing studies seem to suggest that teachers not only hold different attitudes and beliefs about non-White students, but also treat them differently than White students. Generally, current research suggests that

teachers expect Black and Hispanic students to perform less well than White students, and expect Asian students to perform better than White students (e.g., Baron, Tom, & Cooper, 1985; Irvine, 1990; Tenenbaum & Ruck, 2007), although it is unclear whether these different expectations reflect discrimination versus true performance differences (Ferguson, 2003). Teachers also have been shown to unconsciously give varying amounts of positive or negative feedback based on students' race (e.g., Feldman & Orchowsky, 1979; Taylor, 1979). Although these differential attitudes and nonverbal behaviors do not constitute evidence that teachers' grading decisions are themselves biased, they do support the idea that teachers both experience and act upon some amount of racial bias.

Some early experimental studies found differences in teacher expectations and behaviors based on race. In an experimental study (Feldman & Orchowski, 1979), in which both the race and performance of students was controlled, White female students (freshmen in an Introduction to Psychology course) were asked to play the role of a teacher, and to give positive or negative feedback to students. The specific wording of the feedback was controlled; participants were told what to say if an answer was correct or incorrect. Sessions were videotaped to record nonverbal behavior. Although participants were able to correctly give the appropriate verbal feedback to students regardless of race, their nonverbal feedback was perceived as significantly more positive when the learner was White than when the learner was Black. These nonverbal behaviors were also picked up on by new participants who were asked to rate the teacher's pleasure or displeasure with students based solely on visual cues, without being able to hear the words spoken.

This shows that not only do nonverbal behaviors change depending on the race of

the student being evaluated, but that these changes are perceptible to average people who are not psychologists; students most likely are aware of it when a teacher treats them differently. However, interestingly, in this study the evaluations themselves did not actually differ based on the race of the student. Participants rated students as being more successful, and their enjoyment of the teaching exercise, based on student performance, not race. This suggests that evaluations of performance, at least straightforward and objective ones, may be based on actual performance and unaffected by bias. It also suggests that the individuals displaying differing nonverbal behavior may be doing so unconsciously, because participants did not actually express differing levels of pleasure or displeasure with the Black compared to the White students. Because this study was conducted using college students in a controlled setting rather than actual teachers, it is unclear whether these results can be generalized to real teachers in an actual classroom setting. Nevertheless, this study opens up questions as to whether teachers' real feelings toward students may be affected by race, and if they are aware of it.

Another experimental study (Taylor, 1979) assigned trainees in teacher training programs to students to teach in order to determine how they might treat students differently based on their race and ability levels. In this study, there were no real students, but teachers were given a description of the pupil they were supposedly teaching, and were told the pupil was participating in the lesson from behind one-way glass. Although teachers who were assigned Black pupils actually reported higher expectations of these pupils, there was no evidence that these students received better treatment. In fact, teachers were more likely to omit positive feedback after correct responses, and less likely

to make slips of the tongue that would be beneficial to the students, when the students were Black. Interestingly, this study also found an interaction between the effects of race and the effects of student ability, such that teachers were more likely to speak in warm tones and make helpful slips if the student was a high ability Black student, or a low ability White student.

Several reviews of race-related differences in teacher attitudes and behaviors have found some inconsistencies in results, but generally concluded that teachers have higher expectations of White students than Black students. A study by Baron, Tom and Cooper (1985) summarized the existing teacher expectations literature. Specifically, they focused on whether teacher expectations were higher for White students than for non-White students. They were able to locate 16 studies with experimental designs, in which students of different races were randomly assigned to teachers, or college students role-playing as teachers. The race manipulation was accomplished in most studies using either a written description, a photograph, or a video tape. One study used an audio tape to manipulate dialect, and one used actual confederates. Of the 16 studies, 9 of them concluded that teachers had higher expectations of White students. Only one study found that Black students were favored over White students. The remaining 6 studies did not report whether White or Black students were favored. This review is limited in that the studies primarily focused on elementary school students, and the "teachers" were not actually teachers at all in roughly half of the studies, but rather were students who were role-playing as teachers. Also, most of the studies were published prior to 1980, so it is difficult to say what changes in racial climate might have occurred in more recent years. There might also be

differences at the college level in how teachers view their students, given that they perhaps have other facts on which to base their initial impressions. For example, they are aware that their students were capable of completing high school and getting into college.

Irvine (1990) conducted a review of 34 studies completed before 1990 to summarize the findings of studies on teacher perceptions of Black students. Of these studies, 18 used surveys asking teachers about their perceptions of Black students compared to White students in general. They measured attitudes and perceptions regarding ability, personality traits, language ability, behavior, and potential for achievement. In all 18 of these studies, teachers reported having higher expectations of White students. The other 16 studies used more naturalistic classroom settings, measuring teacher expectations of their students either with surveys or by observing their interactions, including eye contact, positive and negative feedback, and other verbal interactions. The results of these studies were more mixed. In 9 of these 16 studies teachers expected more from White students, in 4 studies no difference was found, one study found teachers expected more from Black students, and 2 studies found that teachers expected more from students not from their own racial subgroup. Although the results of these studies were not completely consistent, the majority of them (at least 27 out of 34) found that teachers have higher expectations of White students than Black students.

A more recent study by Tenenbaum and Ruck (2007) examined racial differences in the way teachers interact with their students, comparing European American students to Asian, Black, and Hispanic students. Specifically, they conducted 4 separate meta-analyses looking at expectations (39 samples in which teachers were asked to rate

children from different ethnic groups, including either their own classes or hypothetical children); disciplinary, gifted, and special education referral rates (15 samples, some of which used objective records and some of which asked teachers to answer survey questions about vignettes); positive/neutral speech patterns (11 samples); and negative speech patterns (10 samples). Although teachers did not direct differing amounts of negative speech depending on students' races, they were more likely to direct positive or neutral speech to European American students than to Hispanic or Black students 0.21). The results of the expectations meta-analyses showed teachers had higher expectations of Asian students (d = -0.17) than European American students, and generally more positive expectations of European American students than Black (d = 0.21) or Hispanic students (d = 0.46).

Conclusions. Although the above studies are evidence of differences in the educational experiences of White versus non-White students, there is no clear evidence specifically for or against racial bias in grading practices. Although both experiences of racial discrimination and evidence of differential expectations and treatment based on race are indirect lines of evidence regarding whether racial biases exist, they are not direct evidence that these biases affect grading practices.

Alternative Explanatory Factors

Although psychological theories of racism, along with research in educational settings, suggest the possibility that some form of bias could affect the grades of non-White students, there are a number of factors, other than bias in grading, that could make the grades of non-White students more variable than those of White students. Other

factors that have been found to be related to the race of students might also contribute to differences in grade variability. Although determining if there is racial bias in grading would require further study, it is possible to model and control for some of these other factors, to determine whether they account for differences in grade variability between subgroups. Students' level of preparation for college, upbringing and stability of home life, familiarity with the English language, and academic performance (i.e., college GPA) might all contribute, in different ways, to students of different races receiving more variable grades than White students. If these factors can account for differences in grade variability between races, then bias in grading becomes a less likely explanation. Thus, the present study determines the amount of variance that can be attributed to these non-bias factors.

College Preparation. One factor that might differ systematically between racial subgroups, causing non-White students' grades to be more variable, is differences between subgroups in college preparation (e.g., Young & Kobrin, 2001; An, 2010; Niu & Tienda, 2010). Two ways to operationalize college preparation are to use SAT scores and high school GPA, because students with lower SAT scores and/or high school GPAs are likely to be less prepared for the rigors of college. Being less prepared for college coursework could lead to a number of different factors that would vary over the course of a college career, such as remedial coursework or not taking a full course load initially (Willingham, 1985). Many of the students who enter college, especially those from poorer backgrounds where the available public schooling may have been of lesser quality than for those students from privileged backgrounds, might be underprepared for the rigors of

normal college courses. In these cases, students may choose, or be required, to take remedial courses for the first semester or so in order to gain the necessary skills for more advanced work. Grades early in their curriculum, then, would be for classes that are not at the same difficulty level, and perhaps not graded on the same scale, as other courses. Therefore, grades in earlier courses might be less correlated with grades in later courses. It might also be the case that a student only needs remedial help in some areas, but is able to take some normal courses as well. Once again, the remedial course grades might not be as highly correlated with normal college course grades.

A lack of preparation for college might also lead a student to take fewer courses early on, rather than take on a full college course load all at once. A student who takes fewer than the normal number of hours at one point in the college career, and goes on to take a normal course load (or an overload) at another time, might have grades that are less intercorrelated than a student who takes a steady course load. If a student were to sign up for, say, only one class because of a lack of preparation in the first semester, then that student might find it easier to do well than he or she would have had he or she taken more classes at the same time. After that first semester, maybe the student would decide to take on a full course load. With so many additional demands now on time, there would not be as much time to devote to each class. In addition, the mental faculties required to learn several subjects concurrently are higher than those required to learn only one subject at a time. Therefore, although both semesters would be reflective of a student's ability, they would be less correlated with one another than if the course load had been split evenly between the two semesters, because that would even out the demands on the student's

time.

Black and Hispanic students, on average, have lower educational admissions test scores and lower high school GPAs than Asian and White students (e.g., Young & Kobrin, 2001), and yet a recent study found that they are more likely to apply to selective schools than White students, which also may mean they are less likely to be prepared for these rigorous programs (An, 2010). A nationally representative sample of tenth grade students was surveyed in 2002, with follow up surveys in later years. Regression analyses showed that Asian, Black and Hispanic students had a greater probability of applying to a selective college than comparable White students. Black and Hispanic students are also more likely to come from poorer school districts where they may have received inflated grades (Aud et al., 2010), meaning less prepared students may get in to more selective and academically rigorous institutions. Niu and Tienda (2010), using data from the University of Texas at Austin between 1990 and 2003, evaluated claims that students granted automatic admission based on top 10 percent high school class rank underperform academically relative to lower ranked students who come from highly competitive high schools. Black and Hispanic students who were in the top 10 percent of their graduating high school classes received lower average standardized test scores than White students rated at or below the third decile (i.e., top 30%), but they consistently performed as well or better in grades, first year persistence, and 4 year graduation likelihood.

Socioeconomic Status (SES). Another factor that could affect the amount of variability in a student's grades is the stability of his or her life at home. For a number of reasons, SES might play a role in the level of stability, and the type of life in general, that

a student has at home. For example, a student from a poor family might be more likely to have to work a job in order to support him or herself, and might therefore have to base class schedules and study time on a job. This could affect the variability of course load throughout college, as well as the amount of time the student would be able to put into school work. The requirement to support oneself, or even one's family, while in school would be an important additional source of variance that might be captured in the measure of GPA.

Non-White students often come from lower income households with a lesser amount of stability than White students. For example, Black and Hispanic students are more likely to come from fragile families, which have been linked with lower SES and generally unstable home environments that might make earning consistent grades more difficult (Hummer & Hamilton, 2010). Fragile families include those where a child is born to unmarried parents; this family structure is related to lower SES and living in unsafe neighborhoods. Hummer and Hamilton (2010) compared data from the National Center for Health Statistics (1970-2006) for Black, American Indian, Hispanic, and White individuals, and found that Black (75% in 2006) and Hispanic (50% in 2006) mothers had a greater share of births to unwed mothers than Whites (27%) and Asians (17%). Unmarried new mothers in all groups were more likely to have less than a high school education, have a partner with less than a high school education, and live in or near poverty, than married new mothers in the same group, and these differences were more pronounced for Black, Mexican American, and Mexican immigrant mothers compared to White mothers.

It is also possible that the effects of low SES may be different for non-White students than White students. Young et al. (2011) looked at the relationship between SES, perceived social support, and both intrinsic and extrinsic motivation for Black, Hispanic, and White undergraduate students. This study found that SES and perceived social support were positively associated with both intrinsic and extrinsic motivation for Black students; however, the results were non-significant for White and Hispanic students. Interestingly, perceived social support was measured using 3 subscales, including family support, friend support, and support from a professor. This indicates that, at least for Black students, all of these factors (which may be impacted by race) have the potential to decrease their motivation to succeed in school.

In order to compare Asian American students to students of other races in terms of the factors that affect their likelihood of obtaining a college degree, Vartanian et al. (2007) examined data from the National Education Longitudinal Study (year 2000 data), which had a nationally representative sample of eighth graders (including Asian, White, Black, Hispanic, and American Indian subgroups) in 1988, who were surveyed again in 1990, 1992, 1994, and 2000 (N = 9494). The authors compared the effects of immigrant status, parental educational orientation, SES, and early school achievement on the likelihood of college graduation for Asian Americans in comparison with other racial groups. They found that parents of Asian students were more educated and had higher SES, and Asian students were much more likely to live with both a mother and father, and to have an immigrant parent (76% of Asian respondents had an immigrant parent). Asian students also had higher eighth grade GPA and standardized test scores. Higher SES, residential

stability, and living in a 2 parent household were associated with a greater probability of completing college. Although Asian students were more likely to complete a college degree than were Black, Hispanic, and White students, this increased likelihood was largely accounted for by higher SES, likelihood of having an immigrant parent, and higher parental expectations.

English as Best Language. Another race-related factor that could contribute to variability of grades is the possibility of entering college when English is not one's best language. Writing intensive courses require a thorough mastery of the language, but more mathematically oriented classes might not emphasize the use of language as much. For this reason, one might expect a student for whom English is a second language to have English grades that are less correlated with, say, their math grades than a student for whom English is the primary language spoken. In addition, students who are not strong English speakers might improve their English skills throughout the course of their college career. In these cases, grades in later courses might be less correlated with grades in earlier courses than the grades of a student for whom English skills did not change.

First or second generation immigrants may be less likely to speak English as a first language, and their grades might vary depending on the degree of English fluency required by the course. English speaking ability is an important consideration when comparing the academic achievement of different racial groups because non-White students (Asian and Hispanic in particular) are more likely than White students to speak a language other than English. According to data from the National Center for Education Statistics (NCES), in 2007, 69% of Hispanic elementary/secondary school students and

64% of Asian students spoke a language other than English at home, compared to 6% of both Black and White students. Additionally, 18% of Hispanic and 17% of Asian students had difficulty speaking English, compared with just 1% of both Black and White students (Aud et al., 2010). Although high school GPA and SAT scores have been shown to be predictive of freshman year grades, there are differences in achievement patterns between students for whom English is a first language and students for whom it is not. Zwick and Sklar (2005) used data from the High School and Beyond (HSB) survey, conducted by the NCES, the base-year survey being conducted in 1980 and the latest follow-up in 1992, to determine how well SAT and high school grades predicted freshman GPA and college graduation for Hispanic students whose native language was Spanish compared to Hispanic, Black, and White students whose native language was English. This study was based on the data collected in the fourth follow up survey of the 1980 sophomore cohort, which included 14,825 of the sophomores originally sampled. They found that prediction of freshman GPA for the Hispanic/Spanish group was much weaker than for the other groups, largely because the predictive ability of high school GPA was so low for this group. In comparing Hispanic/Spanish and Hispanic/English groups, they found that freshman GPAs were similar, but the Hispanic/Spanish group did have slightly higher freshman GPAs. Also, a greater proportion of the Hispanic/Spanish group attended selective colleges compared to the Hispanic/English group. Both freshman GPA and graduation were more strongly associated with the academic predictor variables for the Hispanic/English group than the Hispanic/Spanish group. Because Hispanic students are more likely than White students to speak Spanish as a first language, it is important to take

this difference into consideration when looking for race-related differences in academic performance. The finding that high school GPA was less predictive of freshman GPA for Spanish-speaking students is evidence that differences in first language need to be controlled for. English speaking ability is an ability which might easily change over a college career if a student were less familiar with English upon entering college, and these changes over time could be a factor in the difficulty of predicting freshman GPA for these students.

College GPA. Finally, cumulative college GPAs of non-White students differ from those of White students. Recent College Board data from 66 four-year colleges, including 39,440 students, shows that the mean cumulative second-year GPA for White students who entered college in 2006 was 3.12. Asian students' GPAs were slightly higher at 3.18. Black students and Hispanic students had lower GPAs; their cumulative second-year GPAs were 2.72 and 2.96, respectively (College Board, 2012). There are a number of reasons to believe college GPA will be related to variability in grades, mainly because college GPA can be seen as a proxy for academic ability itself; students who have higher GPAs have performed better in their classes and thus have received better grades. In addition to the number of factors that I have already suggested might influence students' grade variability, variability in course difficulty can also cause grade variability. Even within each major field of study, there is natural variability in course difficulty; some subjects are easier to master than others, and some professors have higher expectations than others. For a good student, this variability in course difficulty might not have any impact on the actual variability in grades; a strong student who is able to earn an

A in a difficult course will also be able to earn As in easier courses. However, although weak students might be able to earn As in particularly easy courses, they will most likely earn lower grades in the more difficult courses. Therefore, the natural variability in course difficulty will be more likely to affect grade variability for weak performers (i.e., students with low GPAs) than high performers (i.e., students with high GPAs), making grades more variable for students with lower GPAs.

In addition to students with lower academic ability receiving varying grades due to natural variations in course difficulty within their major field of study, previous research has shown that students differ in course selection based on their perceived ability to earn a good grade (e.g., Berry & Sackett, 2009). This means that many students who begin in difficult majors will fail to do well and then switch to a less difficult major. Students who have lower academic ability are more likely to find the major they begin in to be too difficult, and then switch to an easier major (Willingham, 1985). These students' grades would therefore be more variable, because their cumulative GPAS would include both the lower grades earned in the original difficult major and the higher grades earned after switching majors. Students who perform well from the beginning, on the other hand, will not show this variability in performance.

Finally, it is also the case that there is just less opportunity for high performing students' grades to vary, because the highest grade a student is able to earn is an A. For a student whose average level of performance results in A grades, it is impossible for extraordinary performance to result in a grade that is *higher* than an A; there is a cap on the maximum grade that can be achieved. Therefore, although a student with a B average

might sometimes earn higher than a B and sometimes earn lower than a B, a student with an A average might sometimes do worse, but will never do better. The ceiling effect experienced by high performing students means that these students will have less variable grades than students who are lower performing.

Conclusions. If there are differences between racial subgroups in the variability of grades, it is important to determine which explanatory factors might cause such a difference. Some possible explanatory factors include differences in college preparation, socioeconomic status, English speaking ability, and actual academic performance. Being less prepared for college coursework could lead to a number of different factors that would vary over the course of a college career, and would thus affect that variability of grades. Lower socioeconomic status could also mean that financial challenges across the college career might lead to a student having variable amounts of time to spend on studying. Additionally, English skills for students who learned to speak a different language first, might be more important in some classes than others, and might also improve across the college career, both of which might make grades more variable for these students. Finally, differences between students in academic performance could account for differences between students in grade variability, and stronger academic performers are also likely to be *consistently* strong performers. Because all of these factors have the potential to affect the variability of students' academic performance, it is important to determine whether race differences in these alternative explanatory factors are the cause for race differences in within-persons grade variability.

CHAPTER II

PRESENT STUDY

The present study seeks to determine whether non-White students' grades are more variable than White students' grades. This study first investigates whether the intercorrelation of grades differs based on the race of the student. This contribution expands current knowledge of the psychometric properties of grades, as I am aware of no large sample studies reporting intercorrelation estimates of college grades separately by race. Next, the present study determines if non-White students' grades are in fact more variable than White students' grades, and if these differences in variability (if present) can be accounted for by a number of possible explanatory factors.

Differential Intercorrelations of Grades

The first goal of this study is to determine whether individual student grades are equally intercorrelated for students of different races, to determine whether there is a possibility that racial bias and/or race differences in course load, college GPA, high school GPA, SES, English as best language, and SAT scores might lead to the grades of non-White students being less intercorrelated than those of White students. Because GPA is used as the criterion measure in studies of the effectiveness of various predictors of academic performance used in college admissions systems (e.g., SAT/ACT, HSGPA; Kuncel, Hezlett, & Ones, 2001; Sackett et al., 2009) and because standardized tests have been shown to be less related to academic performance for students of different races (Berry, Clark, & McClure, 2011; Culpepper & Davenport, 2009; Pearson, 1993), it is

important to determine whether there are subgroup differences in the intercorrelations of student grades. The present study calculates the average intercorrelation between grades that students receive throughout their college careers (i.e., what is the correlation between the grades students earn in their first and second college courses, between their first and third college courses, etc.).

The present study postulates that the grades of non-White students (Asian, Black, and Hispanic) will exhibit different properties than the grades of White students. One reason to expect this is that racial bias might affect the grades of non-White students, without affecting the grades of White students. For instance, if there was any bias toward non-White students, this bias would contaminate the grades of these students and act as construct-irrelevant variance. Some instructors might grade non-White students especially harshly, but others might grade the same students leniently. Thus, a given grade for a non-White student would be less indicative of grades received in other courses, since instructors would likely differ in the degree to which their grading policies reflected true performance versus racial bias. Race differences in college preparation, course load, English as best language, SES, and actual college performance could also make the grades of non-White students less intercorrelated. For these reasons, I predict that the grades of non-White students will be less intercorrelated than the grades of White students.

Hypothesis 1a: The average intercorrelation of grades received throughout the college career will be higher for White students than for Asian students.

Hypothesis 1b: The average intercorrelation of grades received throughout the college career will be higher for White students than for Black students.

Hypothesis 1c: The average intercorrelation of grades received throughout the college career will be higher for White students than for Hispanic students.

Race Differences in Standard Deviation of Grades

As mentioned above, there exists virtually no empirical research on racial bias or discrimination in grading, despite (1) the importance of this issue, (2) empirical evidence that teachers can exhibit negative behaviors and expectations toward non-White students, and (3) theoretical reasons for expecting race to play a role in grading. The present study will use a large educational dataset to take the first steps toward remedying this lack of research. Current theories of racism suggest that different professors will tend to grade non-White students either more harshly or more leniently than White students. To the degree that this is true, non-White students would have more variable grades than White students. Therefore, the typical non-White student would have a wider range of grades received than the typical White student, and the average within-person standard deviation of grades for non-White students would be higher. For example, if a White and non-White student both preformed at a "B" level in 3 different classes, the White student would likely receive a B in all 3 classes, but the non-White student's grades might vary depending on the professor teaching each course. Therefore, the non-White student might receive a B from a non-biased professor, a C from a prejudiced professor, and an A from a professor with a bias in favor of non-White students. In this case, the within-person standard deviation of grades for the White student would be 0.0, as there is no variability in grades across the three courses; the non-White students' standard deviation would be 1.0, as this is the standard deviation of 2.0, 3.0, and 4.0. Thus, the racial bias in grading perspective

would hypothesize that:

Hypothesis 2a: The average within-person standard deviation of the distribution of grades received throughout the college career will be greater for Asian students than for White students.

Hypothesis 2b: The average within-person standard deviation of the distribution of grades received throughout the college career will be greater for Black students than for White students.

Hypothesis 2c: The average within-person standard deviation of the distribution of grades received throughout the college career will be greater for Hispanic students than for White students.

However, the existence of race differences in the within-persons standard deviation of grades does not necessarily mean there is bias in grading. There are several other race-related factors outside of the possibility of racial bias that could cause non-White student grades to be more variable than those of White students; therefore, it is important to explore these subgroup differences as possible explanations for differences in grade variability. Because existing research has shown race-related differences in college preparation, college GPA, SES, the likelihood of English being a student's best language (e.g. Young & Kobrin, 2001; An, 2010, Hummer & Hamilton, 2010; Zwick & Sklar, 2005), and because each of these factors could potentially influence the variability of student grades, the present study controls for these differences and factors related to them. The present study controls for the effects of course loads, high school GPA, English as best language, SAT scores, SES, and college GPA on within-person standard deviations

of grades. If the race differences in within-persons standard deviations of grades disappear once these factors are controlled, racial bias in grading becomes a less tenable explanation. If these factors do not account for the race differences in grade variability, racial bias in grading remains a possibility. It is not clear whether these other factors will or will not account for racial differences, so there are no specific hypotheses for this question.

Research Question 1: Do differences between subgroups in course loads, high school GPA, English as best language, SAT scores, SES, and college GPA account for differences between subgroups in the within-person standard deviations of grades?

Finally, as was mentioned previously, in order for racial bias to have an effect on college grades, a number of conditions would need to be met. First, the individuals responsible for student grading would have to hold racial biases that could affect their judgment. Current theories of racial bias suggest this is likely the case; however, there are two other important conditions that must also be met. The first of these is that there must be some amount of room for subjective judgment in the grading process. The second is that the individual responsible for giving the students' grades would need to be aware of each student's race. Without such knowledge, the expression of racial bias would be impossible. It is unlikely that all three conditions will be met in every class a student takes, and in this dataset it was impossible to definitively differentiate small, subjectively graded courses from large, objectively graded ones. However, smaller private schools tend to have smaller class sizes in general. Courses taken later in the college career, also, are

likely to be smaller even in large universities. When class sizes are smaller, it is more likely that (1) the person grading students is aware of students' races, and (2) students are given at least some assignments that leave room for subjective judgment; purely objective assignments are much easier to grade, and thus more likely in large introductory classes. For this reason, if there were racial bias in grading, I predict there will be differences between the pattern of race-related differences seen in public vs. private schools, and in early vs. late year classes. Specifically, I predict that race differences in both grade intercorrelations and within-person standard deviations of grades will be smaller in public schools than in private schools.

Hypothesis 3a: The Asian-White difference in average grade intercorrelations will be smaller in public schools than in private schools.

Hypothesis 3b: The Black-White difference in average grade intercorrelations will be smaller in public schools than in private schools.

Hypothesis 3c: The Hispanic-White difference in average grade intercorrelations will be smaller in public schools than in private schools.

Hypothesis 4a: The Asian-White difference in average within-person standard deviation of grades will be smaller in public schools than in private schools.

Hypothesis 4b: The Black-White difference in average within-person standard deviation of grades will be smaller in public schools than in private schools.

Hypothesis 4c: The Hispanic-White difference in average within-person standard deviation of grades will be smaller in public schools than in private schools.

I also predict that race differences in the within-person standard deviations of

grades will be lowest in students' first year of college, but will increase over time, such that they are largest in students' fourth year of college.

Hypothesis 5a: The Asian-White difference in average within-person standard deviation of grades will increase over time, such that differences will be smallest in the first year and largest in the fourth year.

Hypothesis 5b: The Black-White difference in average within-person standard deviation of grades will increase over time, such that differences will be smallest in the first year and largest in the fourth year.

Hypothesis 5c: The Hispanic-White difference in average within-person standard deviation of grades will increase over time, such that differences will be smallest in the first year and largest in the fourth year.

CHAPTER III

METHODS

Participants

Participants were drawn from a sample provided by the College Board of 167,816 students representing 41 U.S. colleges' entering classes of 1995-1997 that had taken the SAT. Of these students, 150,417 students self-reported their race on the Self Descriptive Questionnaire (SDQ; a questionnaire completed when they took the SAT) as either Asian (N = 19,215), Black (N = 8,918), Hispanic (N = 10,267) or White (N = 111,981); and thus comprised the study sample.

Measures

College Grades. Each college provided the first 60 college grades (reported on a scale of 0.0-4.25) received by each participant, resulting in 4,423,433 grades earned; 639,851 by Asian students, 267,552 by Black students, 298,331 by Hispanic students, and 3,517,699 by White students. Only grades that were included in the final GPA of each student were included; therefore, grades from dropped classes were not included.

Race. Race information was obtained from the SDQ. To compare grade variability for each non-White group to that of White students, three dummy variables were created, such that each non-White group was equal to 0 and the White group was equal to 1.

Course Load. A course load variable was created for each student based on the total number of grades reported for that student.

High School GPA. High school GPA was self-reported on the SDQ on a scale of

0.0-4.0. Although obtaining the actual GPA of students from their high schools would have been ideal, it has been shown that self-reported high school GPA is a valid predictor of future performance, and that self-reported GPA is more consistent with actual GPA for high performing students (Kuncel, Crede, & Thomas, 2005). Because this data set is made up of relatively high performers (i.e., those who were accepted to college), the use of self-report GPA was determined to be appropriate.

English as Best Language. The English as best language variable was based on a self-report measure on the SDQ where students were asked to report whether (1) English was their best language, (2) another language was their best language, or (3) English and another language were about the same. A 0-1 dummy variable was created, with English as best language being 1 and another language being either best language or the same as English as 0.

SAT. The College Board provided SAT-Verbal and SAT-Mathematical scores for each student. These were combined into a unit-weighted composite. In cases where a student had more than one SAT score, the highest score was used.

SES. Three SES variables were obtained from the SDQ: father's education, mother's education, and family income. In order to capture the non-linear relationship between income and education attainment, the natural log of family income was used. This study's measure of SES was an equally weighted composite of these three variables.

Cumulative College GPA. Each college provided students' cumulative grade point averages to the College Board. GPAs ranged from 0.0 – 4.25.

Analyses

Intercorrelation Analyses. Average grade intercorrelations were calculated within each of the 41 schools and within each of the four racial subgroups (Asian, Black, Hispanic, and White), in order to determine whether there were race differences in average grade intercorrelations. To calculate these average grade intercorrelations, I first calculated the pairwise correlations between student grades (i.e., the correlation between the first and second grade students received in college, the correlation between the first and third grade, etc.), and then averaged all of these pairwise correlations to get the average grade intercorrelation within each school overall, as well as within each racial subgroup within each school. Using Hunter and Schmidt's (2004) bare-bones meta-analysis method, I then calculated sample size-weighted average grade intercorrelations overall and for each racial subgroup. Finally, in order to compare results for public vs. private schools, the sample size-weighted mean grade intercorrelations for each race were also calculated for public and private schools separately.

Race Differences in Standard Deviation of Grades. In addition to race differences in grade intercorrelations, it is also possible that the grades of non-White students are more variable than the grades of White students. Therefore, I also calculated the standard deviation of grades across the college career for each student to determine whether the average standard deviation differed by race. The within-person standard deviation (SD) of grades throughout the college career was first calculated for each student (e.g., if Student A took three college courses and received grades of 3.0, 3.3, and

4.0; that student's standard deviation would be .51). These within-person SDs of grades were then averaged within each racial subgroup to estimate the average variability of within-person grades for Asians, Blacks, Hispanics and Whites. Average within-person SDs of grades were also calculated for each race for each of years 1-4, as well as separately for public and private schools.

The SDs of grades were also regressed, in a series of hierarchical regressions, on a number of variables that might account for racial differences in the variability of grades (i.e., the "alternative explanatory factors"). These regressions were carried out within schools, and then the regression weights were sample-size-weighted and averaged to obtain final regression weights. Three race dummy variables (Asian = 0, White = 1; Black = 0, White = 1; and Hispanic = 0, White = 1) were created for use in these analyses. In Step 1 of the first set of regressions, SDs of grades were regressed on one of the race dummy variables. The resulting unstandardized regression coefficients represented the mean differences between each non-White subgroup and the White subgroup in SDs of grades (since the coefficient was equal to the mean increase in *SD* of grades resulting from a one-unit increase [0 to 1] in race). Because the intercept in these regressions was equal to the White mean SD, and the unstandardized regression coefficients were equal to the difference between the White subgroup and the non-White subgroup, it was also possible to determine each subgroup's mean SD, and thus compare all groups to one another. In the second step, course load, cumulative college GPA, high school GPA, English as best language, SAT, and SES were added to the regression equation. Any change in the dummy variable's unstandardized regression coefficient represented the effect of

accounting for these additional variables on racial differences in variability of grades.

Four further sets of hierarchical regressions included the same predictors as the above regressions, but used only grades from one of the first four years in college as criteria. In the first, within-person SDs of grades received only in the first year of college were regressed on each race dummy variable, and then on the additional explanatory variables. The next three sets included grades for Year 2, Year 3, and Year 4, respectively.

CHAPTER IV

RESULTS

Intercorrelations

In partial support of Hypotheses 1b and 1c, the average intercorrelation between grades was higher for White students than Black or Hispanic students. However, Hypothesis 1a was not supported; the average intercorrelation of grades for Asian students was actually slightly higher than that for White students. See Table 1 for a summary of intercorrelations of grades. For Asian and White students, the average intercorrelations of grades were .255 (*SD* = .043) and .253 (*SD* = .035), respectively, but for Black and Hispanic students the intercorrelations of grades were .198 (*SD* = .037) and .211(*SD* = .044), respectively. Confidence intervals for Asian and White students versus Black and Hispanic students did not overlap.

In answer to Hypotheses 3a-c, this pattern of results was similar in both public and private schools, with one small difference. In public schools, the pattern of results was similar to the overall pattern. That is, Asian students' grades were the most intercorrelated (*M* = .260, *SD* = .041), followed closely by White students. Hispanic (*M* = .209, *SD* = .039) and Black (*M* = .195, *SD* = .033) students had less intercorrelated grades. The pattern of results in private schools, however, appeared slightly different for Asian students, whose average grade intercorrelation was .225 (*SD* = .045) in the private schools, closer to the intercorrelations of Hispanic (*M* = .225, *SD* = .062) and Black (*M* = .212, *SD* = .158) students than to the intercorrelation for White (*M* = .250, *SD* = .036)

Table 1

Average Sample-Size-Weighted Intercorrelations of Grades by Race for All Schools, Public Schools, and Private Schools

	Race	N	*k*	*r*	SD_r	SD_{Res}	%Var	CI_L	CI_U
All Schools	All Students	149,87	41	.253	.034	.030	21.00	.243	.264
	Asian	19,210	41	.255	.043	.003	99.62	.241	.268
	Black	8,884	41	.198	.037	0	100	.178	.218
	Hispanic	10,231	41	.211	.044	0	100	.192	.229
	White	111,55	41	.253	.035	.030	26.31	.242	.264
Public	All Students	123,65	26	.254	.034	.031	15.87	.241	.267
	Asian	15,886	26	.260	.041	.015	86.69	.245	.276
	Black	7,654	26	.195	.033	0	100	.174	.217
	Hispanic	8,551	26	.209	.039	0	100	.188	.228
	White	91,567	26	.253	.035	.031	2.58	.240	.267
Private	All Students	26,221	15	.252	.030	.023	48.35	.234	.267
	Asian	3,324	15	.225	.045	0	100	.194	.258
	Black	1,230	15	.212	.058	0	100	.157	.265
	Hispanic	1,680	15	.225	.062	0	100	.179	.270
	White	19,987	15	.250	.036	.025	51.13	.233	.269

Note – The above intercorrelations were arrived at by (1) calculating all pairwise intercorrelations between the first 60 student grades within each college, both overall and within each race, (2) averaging these intercorrelation, both overall and within each race, for each school, (3) weighting each within-college intercorrelation by its sample size, and (4) averaging across these sample-size-weighted intercorrelations. *r* = mean sample-size-weighted correlation; SD_r = sample-size-weighted observed standard deviation of correlations; SD_{Res}= standard deviation of intercorrelations after removing sampling error variance; %Var = percentage of variance attributable to sampling error; CI_L and CI_U = lower and upper bounds, respectively, of the 95% confidence interval around the corrected mean correlation.

students. However, although the mean intercorrelations for Asians at public schools and private schools were different, the 95% confidence interval for Asians at public schools (.245 - .276) overlapped with the 95% confidence interval around the mean intercorrelation for Asians at private schools (.194 - .258), suggesting that public vs.

private school was not a substantial moderator. Thus, Hypotheses 3a – 3c were not supported.

Grade Variability

In support of Hypotheses 2a-c, the grades of each group of non-White students were more variable than the grades of White students. The mean SDs for each subgroup were calculated within each school, and then the sample-size-weighted average *SD* of grades for all schools was calculated. See Table 2 for the results of these analyses. Black students had the highest standard deviation of grades (M = .741, SD = .328), followed by Hispanic students (M = .700, SD = .315), Asian students (M = .651, SD = .295), and finally White students (M = .617, SD = .282). The largest mean difference in grade SDs was between Black students and White students (d = 0.435).

Table 2

Average Sample-Size-Weighted Standard Deviations of Grades by Race for All Schools, Public Schools, and Private Schools

	All Schools		Public Schools		Private Schools	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
All Students	.636	.288	.634	.310	.650	.142
Asian	.651	.295	.647	.31i8	.667	.148
Black	.741	.328	.734	.354	.776	.164
Hispanic	.700	.315	.696	.340	.720	.148
White	.617	.282	.615	.304	.628	.141

Note – The above SDs were arrived at by (1) calculating SDs of grades for each student, (2) calculating the average *SD* of grades, both overall and within each race, for each school, (3) weighting each within-college *SD* by its sample size, and (4) averaging across these sample-size-weighted SDs.

In answer to Hypotheses 4a-c, the pattern of results was similar for both public and private schools. In public schools, Black students had the highest standard deviation of grades (M = .734, SD = .354), followed by Hispanic students (M = .696, SD = .340), Asian students (M = .647, SD = .318), and finally White students (M = .615, SD = .304). The pattern was the same in private schools, where Black students had the highest standard deviation of grades (M = .776, SD = .164), followed by Hispanic students (M = .720, SD = .148), Asian students (M = .667, SD = .148), and finally White students (M = .628, SD = .141). Thus, Hypotheses 4a-c were not supported.

Alternative Explanatory Factors

The lower intercorrelations and greater within-person standard deviations of grades for non-Whites are consistent with a source of additional variance in the grades of non-White students. Therefore, additional regression analyses were run to determine whether these differences could be accounted for by the alternative explanatory factors. The SDs of grades were regressed first on only the non-White-White dummy variables, then on the possible explanatory variables: course load, high school GPA, English as best language, SES, SAT scores, and college GPA (Tables 3-5). Descriptives and correlations between alternative explanatory factors are shown in Appendix A.

In answer to Research Question 1, for each minority group, differences in SDs of grades could be almost entirely attributed to subgroup differences in these alternative explanatory variables. For Asian students, when these additional variables were entered into the regression equation, the Asian-White difference in *SD* of grades decreased from .034 to .012, suggesting that 66% of the Asian-White difference in *SD* of grades was

Table 3

Asian-White Standard Deviation of Grades Difference when Controlling for Course Load, College GPA, High School GPA, English as Best Language, SAT Scores, and SES

	All Years	Year 1	Year 2	Year 3	Year 4
Predictors					
Step 1					
Intercept	.657	.617	.600	.573	.548
Racial/Ethnic Subgroup	-.034	-.019	-.037	-.051	-.059
Step 2					
Intercept	1.435	1.465	1.393	1.328	1.272
Racial/Ethnic Subgroup	-.012	.000	-.005	-.017	-.021
Course Load	.002	.001	.003	.004	.004
College GPA-CUM	-.265	-.215	-.285	-.335	-.357
High School GPA	-.015	-.029	-.006	.010	.007
English as Best Language	-.007	.003	-.009	-.010	-.011
SAT	-.002	.000	.000	.000	.000
SES	-.004	.000	-.002	-.004	-.003

Note – The above regression weights were arrived at by (1) regressing SDs of grades on the independent variables within each of the 41 colleges, (2) weighting each within-college regression weight by its sample size, and (3) averaging across these sample-size-weighted regression weights.
Racial/Ethnic Subgroup = Asian (0), White (1)
SAT has been scaled such that the regression coefficients for SAT represent the change in *SD* of grades for every 100 points of SAT.

Table 4

Black-White Standard Deviation of Grades Difference when Controlling for Course Load, College GPA, High School GPA, English as Best Language, SAT Scores, and SES

	All Years	Year 1	Year 2	Year 3	Year 4
Predictors					
Step 1					
Intercept	.759	.718	.702	.671	.641
Racial/Ethnic Subgroup	-.122	-.105	-.126	-.135	-.141
Step 2					
Intercept	1.410	1.446	1.381	1.304	1.265
Racial/Ethnic Subgroup	-.003	.012	-.002	-.014	-.022
Course Load	.002	.001	.003	.004	.004
College GPA-CUM	-.256	-.206	-.279	-.336	-.360
High School GPA	-.016	-.030	-.006	.011	.007
English as Best Language	-.001	.003	-.009	-.001	-.008
SAT	-.002	.000	.000	.000	.000
SES	-.004	-.001	-.001	-.003	-.001

Note – The above regression weights were arrived at by (1) regressing SDs of grades on the independent variables within each of the 41 colleges, (2) weighting each within-college regression weight by its sample size, and (3) averaging across these sample-size-weighted regression weights.
Racial/Ethnic Subgroup = Black (0), White (1)
SAT has been scaled such that the regression coefficients for SAT represent the change in *SD* of grades for every 100 points of SAT.

Table 5

Hispanic-White Standard Deviation of Grades Difference when Controlling for Course Load, College GPA, High School GPA, English as Best Language, SAT Scores, and SES

	All Years	Year 1	Year 2	Year 3	Year 4
Predictors					
Step 1					
Intercept	.709	.676	.634	.596	.559
Racial/Ethnic Subgroup	-.083	-.074	-.069	-.069	-.067
Step 2					
Intercept	1.418	1.455	1.374	1.307	1.236
Racial/Ethnic Subgroup	-.017	-.012	-.007	-.013	-.013
Course Load	.002	.001	.003	.004	.004
College GPA-CUM	-.255	-.206	-.278	-.332	-.354
High School GPA	-.017	-.029	-.006	.010	.008
English as Best Language	-.005	.004	-.014	-.008	-.009
SAT	-.002	.000	.000	.000	.000
SES	-.005	-.001	-.002	-.004	-.002

Note – The above regression weights were arrived at by (1) regressing SDs of grades on the independent variables within each of the 41 colleges, (2) weighting each within-college regression weight by its sample size, and (3) averaging across these sample-size-weighted regression weights.
Racial/Ethnic Subgroup = Hispanic (0), White (1)
SAT has been scaled such that the regression coefficients for SAT represent the change in *SD* of grades for every 100 points of SAT.

attributable to differences between Asians and Whites in course load, high school GPA, English as best language, SES, SAT scores, and college GPA. For Black students, adding the additional variables into the regression equation resulted in a decrease in the Black-White difference in *SD* from .122 to .003, suggesting that 97% of the Black-White difference in the *SD* of grades was attributable to differences between Blacks and Whites

in these variables. For Hispanics, adding the additional variables into the regression equation resulted in a decrease in the Hispanic-White difference in *SD* from .083 to .017, suggesting that 79% of the Hispanic-White difference in the *SD* of grades was attributable to differences between Hispanics and Whites in these variables.

Of the six alternative explanatory factors controlled for, the majority of the differences between students in grade standard deviations could be attributed to academic performance, as measured by cumulative college GPA. The cumulative GPA regression weights for Asians and Whites (b = -.265), Blacks and Whites (b = -.256), and Hispanics and Whites (b = -.255) all indicated that the average *SD* of grades decreased by just over one quarter of a grade point for each grade point increase in college GPA (e.g., the difference between a 2.0 and a 3.0). Course load (b = .002 for all groups), high school GPA (b < .020 for all groups), English as best language (b < .010 for all groups), SAT (b = .002 for all groups), and SES (b < .010 for all groups) each accounted for only a very small amount of unique variance in average grade SDs, indicating that mean differences in college GPA were the most important factors in race differences in grade variability. This is supported by the actual race differences found in mean GPA in this study (see Table 6). Asian students had the highest mean GPA (M = 2.961, SD = .652), followed by White students (M = 2.864, SD = .731), Hispanic students (M = 2.626, SD = .758), and Black students (M = 2.489, SD = .737).

Variability by Year

Next, to test Hypotheses 5a-c, I carried out the previously described regression analyses four more times, once each including only grades from the first year, the second

Table 6

Average Cumulative GPAs by Race

Race	N	Mean	*SD*
Asian	19,210	2.961	.652
Black	8,884	2.489	.737
Hispanic	10,231	2.626	.758
White	111,554	2.864	.731

year, the third year, and the fourth year. Thus, I was able to determine whether the effects of race on SDs of grades might increase across the college career, as well as whether the amount of variability left unaccounted for by alternative explanatory factors also increased (Tables 3-5).

For Asian students (Table 3) the unstandardized race dummy coefficient increased each year, from .019 (*SD* = .040) in Year 1, to .037 (*SD* = 0.039) in Year 2, .051 (*SD* = .053) in Year 3, and .059 (*SD* = 0.057) in Year 4, meaning that Whites' mean *SD* was .019 less than Asians in Year 1, with the difference slightly but steadily increasing until Year 4 when it was .059. The pattern of results also showed that the proportion of the race difference in grade variability that could be accounted for by alternative factors was smaller each year. The unstandardized race coefficient increased from 0 (*SD* = .027) in Year 1, to .005 (*SD* = .029) in Year 2, .017 (*SD* = .044) in Year 3, and .021 (*SD* = .046) in Year 4. Thus, Hypothesis 5a was weakly supported; the Asian-White difference in grade variability left unaccounted for did increase, but the unexplained difference in Year 4 was still quite small.

For Black students (Table 4), the unstandardized race dummy coefficient increased from .105 (*SD* = .072) in Year 1, to .126 (*SD* = .083) in Year 2, .135 (*SD* = .075) in Year 3 and .141 (*SD* = .076) in Year 4. This means that the Whites' mean *SD* was .105 less than Blacks in Year 1, with the difference slightly but steadily increasing until Year 4 when it was .141. After controlling for alternative explanatory factors, the Black *SD* was actually slightly smaller than the White *SD* in Year 1, for a difference of .012 (*SD* = .036). For the remaining years, the difference was in the hypothesized direction, with the Black *SD* larger than the White SD. In Year 2, the difference was very small at .002 (*SD* = .038); at Year 3 it was .014 (*SD* = .032); and at Year 4 it was .022 (*SD* = .036). Thus, Hypothesis 5b was also only weakly supported; the Black-White difference in grade variability left unaccounted for did increase, but the unexplained difference in Year 4 was still quite small.

For Hispanic students (Table 5) unstandardized race dummy coefficients did not increase over the years. They decreased from 0.074 (*SD* = 0.057) in Year 1, to 0.069 (*SD* = 0.06) in Year 2, and then stayed constant at 0.069 (*SD* = 0.059) in Year 3, and 0.067 (*SD* = 0.059) in Year 4. So, for Hispanic students, the variability differences did not increase but neither did they disappear over time. The pattern of results also showed that the Hispanic-White difference in grade variability could be almost entirely accounted for by alternative factors each year. The unstandardized race coefficient decreased from .012 (*SD* = .037) in Year 1, to .007 (*SD* = .04) in Year 2, then increased again to .013 (*SD* = .048) in Year 3, and remained stable at .013 (*SD* = .037) in Year 4. Thus, Hypothesis 5c was not supported; the Hispanic-White difference in grade variability left unaccounted for by

alternative explanatory factors did not increase over time.

CHAPTER V

DISCUSSION AND CONCLUSIONS

In this study, I used a large educational dataset to investigate whether there might be race differences in the psychometric properties of college grades. Because existing theories of racial bias suggest that professors might grade non-White students more harshly or more leniently than White students depending on the situation, I hypothesized that the average intercorrelations of grades for Asian, Black, and Hispanic students would be lower than those for White students, and that the average within-person SDs of grades for these non-White groups would be greater than those for Whites. Finding that there were indeed race differences in the intercorrelations and SDs of grades, I further asked whether these differences could be accounted for by race differences in course load, high school GPA, English as best language, SAT scores, SES, and college GPA. Finally, I determined whether the pattern of results found would be different across all 4 years of college or whether the race differences found and/or the amount that could be accounted for by alternative explanatory factors would increase over time.

Summary of Findings

First, in order to determine whether there were race differences in the psychometric properties of college grades, I compared the average grade intercorrelations and standard deviations for four different races (Asian, Black, Hispanic, and White). I expected to find that non-White students' grades would be less intercorrelated, and have higher standard deviations, than White students' grades, and my findings were mostly

consistent with my hypotheses. One exception was that Asian students' mean grade intercorrelation was similar to White students'. Also, contrary to my hypotheses, there were no differences between public and private schools in the patterns of average grade intercorrelations or standard deviations of grades; all non-White races had higher grade standard deviations and lower grade intercorrelations than White students in both public and private schools. Asian students' mean intercorrelation was lower, similar to Black and Hispanic students, in private schools; however, overlapping confidence intervals showed that public vs. private schools was not a significant moderator in these results. This means that non-White students' grades do have different psychometric properties than White students' grades; however, the lack of difference between public and private schools shows that any differences between these schools (e.g., possible smaller class sizes, more subjectivity in grading) does not account for these differences.

The next step, then, was to attempt to determine why exactly these properties were different, by controlling for some of the factors that could possibly cause race differences in grade standard deviations (i.e., course load, college GPA, high school GPA, English as best language, SAT scores, and SES). I found that race differences in these alternative explanatory factors did almost entirely account for differences between races in average grade standard deviations; after accounting for these factors, differences between White students and each of the three non-White groups decreased to less than .02 grade points. This suggests that racial bias, as defined earlier in this thesis, is an unlikely explanation for differences between races in grade variability; virtually all of the race differences in grade variability can be accounted for by other non-bias factors.

The unstandardized regression coefficients for each of the alternative explanatory factors indicates the grade *SD* change due to a one unit change in each predictor, holding all other predictors constant; interpreting these coefficients, then, gives further insight into the importance of each of these variables for predicting average grade SDs. Course load, English as best language, SAT, and SES accounted for a very small amount of unique variance in average grade SDs. For course load, a one unit change meant the addition of one class (e.g., the difference between taking 25 classes throughout college and taking 26 classes); English as best language was a dummy-coded variable, meaning a one-unit increase was the difference between speaking English as a best language, or not; for SAT, a one-unit increase was a 100 point difference in SAT score (e.g., the difference between a 1100 and a 1200); for SES, each one-unit difference corresponded roughly to one standard deviation unit difference in the combined scores for fathers' education, mothers' education, and family income. The units for each of these variables, then, are sizeable; however, a one-unit change in each of them, holding all others constant, led to less than $1/100^{th}$ of a point difference in the standard deviation of student grades. A one unit increase in high school GPA led to only a slightly larger difference, with one full grade point decrease (e.g., the difference between a 3.0 and a 2.0) leading to less than .02 points increase in the students' average standard deviation of grades.

The main factor driving the differences between students in grade standard deviations turned out to be academic performance, as measured by cumulative college GPA; the average *SD* of grades decreased by just over one quarter of a grade point for each grade point increase in college GPA (e.g., the difference between a 2.0 and a 3.0). Because

most students' grades will fall within about 2 SDs of their average grade, this means that the distribution of grades for a student with a 2.0 GPA will be more than half a letter grade wider than a student with a 3.0 GPA on both the positive and negative sides of the distribution, making their range of grades received almost an entire grade point wider. Because the regression analyses held constant race and the other alternative explanatory factors, this result can be interpreted as meaning that it is not race, per se, or any of the other alternative factors that cause differences between students in SDs of grades. Rather, it is mean differences in college GPA; students with lower GPAs have more variable grades. The race difference in grade SDs only occurred because of race differences in college GPA.

This is supported by the actual race differences found in mean GPA in this study; White students had the highest mean GPA (M = 2.869, SD = .308), followed by Asian students (M = 2.674, SD = .319), Hispanic students (M = 2.534, SD = .331), and Black students (M = 2.456, SD = .318). There are three main reasons why these differences in cumulative college GPA would have caused differences in grade variability. First, low performers are likely to have more variable grades, because if they do poorly in one class it is likely they will sort themselves into easier classes (Berry & Sackett, 2009; Willingham, 1985); their cumulative GPAs would include both the lower grades earned in the difficult classes and the higher grades earned after choosing easier classes. High performers, on the other hand, will not be as likely to make a conscious effort to choose easier classes (Willingham, 1985). High performers are also less likely to be affected by differences in course difficulty; a low performer might get a good grade in an easy class

and a lower grade in a more difficult class, while a high performer who does well in a hard class will likely also do well in an easy class. Lastly, there is less room for variability in the grades of high performers. While a B student might sometimes do better than average and sometimes do worse, an A student cannot receive higher than an A grade; their grades might sometimes be lower, but they will never be higher.

The results demonstrating college GPA accounts for race differences in grade variability also suggests that race differences in mean GPA may have been a driving factor in race differences in grade intercorrelations as well. However, this was impossible to test directly as there was not a grade intercorrelation for each student that could be regressed on college GPAs or the other alternative explanatory factors.

Finally, I also carried out analyses within each of the first four years of the college career to explore the possibility that grade variability differences might increase over time. I found that for both Asian and Black students, the race differences in SDs of grades did increase slightly over time. For Hispanic students, however, the variability differences remained relatively constant. I also found that the amount of variability that could be accounted for by alternative explanatory factors did decrease percentage-wise across the four years of college; for Asian students, the percentage of variance accounted for decreased from 100% in the freshman year to 64% in the senior year, and for Black students, it decreased from 100% in the freshman year, to 84% in the senior year. For Hispanic students, the amount of percentage of variance accounted for only decreased slightly, from 84% in the freshman year to 81% in the senior year. Despite the pattern of change over time for Black and Hispanic students, however, it is important to note that

race differences in grade SDs left unaccounted for by alternative explanatory factors only increased from very small to slightly less small. Therefore, although there is a small difference between White and non-White students left unaccounted for by factors controlled for in this study, it is unlikely that racial bias (as understood according to current theory) is playing a substantial and systematic role in student grading.

Practical Implications

Subgroup differences in grade variability in the present study are sizable. For example, Blacks' *SD* of grades throughout the college career is about .122 grade points larger than Whites' (d = 0.435). Most of a student's grades will fall within about two SDs of their average grade, meaning the distribution of grades throughout the college career for the average Black student will be about .244 grade points (almost 1/4 of a letter grade) wider on both the positive and negative sides of the distribution. This difference, however, was smaller for Asian and Hispanic students, but the grades of Hispanic students were still .083 grade points more variable than those of Black students, meaning that the distribution of their grades for the average Hispanic student will be about .166 grade points wider than the average White student. The difference was smallest for Asian students, whose grades were .034 grade points more variable than White students.

Because college grades are used both for the validation of selection instruments such as the SAT, and for high-stakes decisions like scholarship awards, internship selection, and even job selection after college, it is important that race differences in the psychometric properties of college grades and the reasons for those differences be as completely understood as possible. One possible explanation initially put forth by this

thesis is the possibility that there is professor bias in grading. The results of this study suggest such bias, based on current theories of racism, is unlikely. Instead, I present evidence of other factors at play that make the grades of non-White students more variable. Specifically, race differences in academic performance, as measured by cumulative college GPA, was found to be the main driver behind race differences in grade variability. Race differences in course load, high school GPA, English as best language, SES, and SAT scores played a small role as well. This means that, although there are race differences in grade variability, these differences are likely due to actual differences in performance, rather than professor bias in grading.

Limitations and Directions for Future Research

I was able to account for the vast majority of race differences in grade variability by controlling for course load, high school GPA, English as best language, SES, SAT scores, and college GPA. However, a number of issues in the present study warrant discussion and suggest further directions for future research. First, although one of the strengths of the present study was its use of data from actual operational settings, the use of preexisting data made experimental control impossible. This means that, although race differences in grade variability appear to be driven primarily by differences in GPA, and the patterns I found were not consistent with my interpretation of current theories of racial bias, I could not definitively identify the causes of any patterns of race-related differences. This is the classic tradeoff inherent in field versus laboratory research, and highlights the need for experimental work dealing with racial differences in academic performance, as well as the possibility of bias or discrimination in grading.

Second, data were not available for a number of additional possible influences on student grades and academic performance in general. For example, factors such as immigrant status, parental job loss, family structure, and adjustment to college have been shown to covary with race, and are predictive of academic performance (e.g. Kalil & Whitman, 2011; Vartanian et al., 2007; Yazedjian et al., 2009). Non-White student performance may also be impacted by self-fulfilling prophesies, stigma consciousness, and stereotype threat (e.g., Guyll et al., 2010). There is a possibility, then, that racial stereotypes may be contributing to differences in student grades without the grading process itself being biased. This study was not able to capture these and other more complex relationships that may be contributing to racial differences. This again highlights the need for further research in this area.

Third, in this study I was not able to take into account a number of characteristics of the courses themselves that undoubtedly impact students' grades. For example, I did not have data on students' majors, or on which courses were directly related to students' major areas. For that matter, this data did not include information on the subject and difficulty of each course. Perhaps most importantly, I did not have information regarding class size or level of subjectivity in the grading for each course. The effects of bias are only possible when (1) there is room for subjectivity in grading decisions, and (2) the person making the grading decision is aware of the race of the student. In this study, I attempted to use the comparison between public and private schools, as well as the comparison between early and later years of college, as a way to determine whether there might be differences in grade variability when students had more face-to-face contact with

their professors; however, these were rough proxies, and I did not find any large or clear patterns of differences between public/private schools or early/late years of college. There were, however, some small differences that could indicate a need for future controlled research on these topics. For example, for Asian and Black students, the differences in grade variability, as well as the proportion of that difference left unaccounted for by the explanatory factors included in this study, increased slightly but steadily across the four years of college. These small differences are not enough to draw definitive conclusions about the effects class size might have on student grading, but they do suggest some possibilities for future research.

Conclusion

The empirical study of racial/ethnic bias in performance ratings has been a common topic of research in two major settings of importance to Industrial/Organizational Psychology: employment and military settings. The present study stands as a first attempt to extend this line of research to a third main setting of importance to our field: post-secondary educational settings. The findings of the present study do not directly refute the existence of bias; however, they do not support the hypotheses I derived from current theories of racial bias. Future experimental research should attempt to control for numerous factors that this study was unable to control for. The results of the present study suggest that, if there is bias in grading, it is not expressing itself by increasing the variability of non-White students' grades. Although Asian, Black, and Hispanic students did have more variable grades than White students, these differences were almost entirely accounted for by race differences in course load, high school GPA, English as best

language, SES, SAT scores, and cumulative college GPA.

REFERENCES

An, B. P. (2010). The relations between race, family characteristics, and where students apply to college. *Social Science Research, 39,* 310-323. doi:10.1016/j.ssresearch.2009.08.003

Aud, S., Fox, M., & KewelRamani, A. (2010). *Status and trends in the education of racial and ethnic groups* (NCES 2010-015). Washington, DC: Government Printing Office. Retrieved from U.S. Department of Education, National Center for Education Statistics website: http://nces.ed.gov/pubs2010/2010015.pdf

Baltes, B. B., Bauer, C. B., & Frensch, P. A. (2007). Does a structured free recall intervention reduce the effect of stereotypes on performance ratings and by what mechanism? *Journal of Applied Psychology, 92,* 151-164.

Baron, R. M., Tom, D. Y. H., & Cooper, H. M. (1985). Social class, race and teacher expectations. In J. B. Dusek (Ed.), *Teacher Expectancies* (pp. 251-269). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Berry, C. M., Clark, M. A., & McClure, T. K. (2011). Racial/ethnic differences in the criterion-related validity of cognitive ability tests: A qualitative and quantitative review. *Journal of Applied Psychology, 96,* 881–906.

Berry, C. M., & Sackett, P. R. (2009) Individual differences in course choice result in underestimation of the validity of college admissions systems. *Psychological Science, 20,* 822-830.

Cascio W. F., & Aguinis, H. (2011). *Applied psychology in human resource management*

(7th ed.). Upper Saddle River, NJ: Prentice Hall.

College Board. (2012). *The validity of the SAT for predicting cumulative grade point average by college major* (Research Report 2012-6). Retrieved from http://research.collegeboard.org/publications/validity-sat-predicting-cumulative-grade-point-average-college-major

Culpepper, S. A., & Davenport, E. C. (2009). Assessing differential prediction of college grades by race/ethnicity with multilevel model. *Journal of Educational Measurement, 46*, 220-242. doi:10.1111/j.1745-3984.2009.00079.x

Feagin, J. R. (1992). The continuing significance of racism: Discrimination against Black students in White colleges. *Journal of Black Studies, 22*, 546-578. doi:10.1177/002193479202200407

Feagin, J. R., Vera, H., & Imani, N. (1996*). The agony of education: Black students at White colleges and universities*. New York: Routledge.

Feldman, R. S., & Orchowsky, S. (1979). Race and performance of students as determinants of teacher nonverbal behavior. *Contemporary Educational Psychology, 4*, 324-333.

Ferguson, R. F. (2003). Teachers' perceptions and expectations and the Black- White test score gap. *Urban Education, 38*, 460-507.

Fiedler, A. M. (2001). Adverse impact on Hispanic job applicants during assessment center evaluations. *Hispanic Journal of Behavioral Sciences, 23,* 102-110. doi:10.1177/0739986301231007

Fischer, M. J. (2010). A longitudinal examination of the role of stereotype threat and racial

climate on college outcomes for minorities at elite institutions. *Social Psychology of Education, 13,* 19-40. doi:10.1007/s11218-009-9105-3

Gaertner, S. L., & Dovidio, J. F. (2005). Understanding and addressing contemporary racism: From aversive racism to the common ingroup identity model. *Journal of Social Issues, 61,* 615-639.

Greenhaus, J. H., & Parasuraman, S. (1993). Job performance attributions and career advancement prospects: An examination of gender and race effects. *Organizational Behavior and Human Decision Processes, 55,* 273-297. doi:10.1006/obhd.1993.1034

Greenhaus, J. H., Parasuraman, S., & Wormley, W. M. (1990). Effects of race on organizational experiences, job performance evaluations, and career outcomes. *Academy of Management Journal, 33,* 64-86. doi:10.2307/256352

Guyll, M., Madon, S., Prieto, L., & Scherr, K. C. (2010). The potential roles of self-fulfilling prophecies, stigma consciousness, and stereotype threat in linking Latino/a ethnicity and educational outcomes. *Journal of Social Issues, 66,* 113-113-130. doi:10.1111/j.1540-4560.2009.01636.x

Hewstone, M. (1990). The 'ultimate attribution error'? A review of the literature on intergroup causal attribution. *European Journal of Social Psychology, 20,* 311-335.

Hummer, R. A., & Hamilton, E. R. (2010). Race and ethnicity in fragile families. *The Future of Children, 20,* 113-113-131. doi:10.1353/foc.2010.0003

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and*

bias in research findings (2nd ed.). Thousand Oaks, CA: Sage.

Irvine, J. J. (1990). *Black students and school failure: Policies, practices, and prescriptions*. Westport, CT: Greenwood Press, Inc.

Kalil, A., & Wightman, P. (2011). Parental job loss and children's educational attainment in Black and White middle class families. *Social Science Quarterly, 92*, 57-78. doi:10.1111/j.1540-6237.2011.00757.x

Kao, G. (2000). Group images and possible selves among adolescents: Linking stereotypes to expectations by race and ethnicity. *Sociological Forum, 15*, 407-430. doi:10.1023/A:1007572209544

Katz, I., Glass, D. C., & Cohen, S. (1973). Ambivalence, guilt, and the scapegoating of minority group victims. *Journal of Experimental Social Psychology, 9*, 423-436. doi:10.1016/S0022-1031(73)80006-X

Katz, I., Wackenhut, J., & Hass, R. G. (1986). Racial ambivalence, value duality, and behavior. In Dovidio J. F., Gaertner S. L. (Eds.), *Racial ambivalence, value duality, and behavior*. San Diego, CA: Academic Press.

Kraiger, K., & Ford, J. K. (1985). A meta-analysis of ratee race effects in performance ratings. *Journal of Applied Psychology, 70*, 56-65.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the Graduate Record Examinations: Implications for graduate student selection and performance. *Psychological Bulletin, 127*, 162–181. doi:10.1037/0033-2909.127.1.162

Kuncel, N. R., & Klieger, D. M. (2007). Application patterns when applicants know the

odds: Implications for selection research and practice. *Journal of Applied Psychology, 92*, 586-593.

Landau, J. (1995). The relationship of race and gender to managers ratings of promotion potential. *Journal of Organizational Behavior, 16*, 391-400. doi:10.1002/job.4030160409

Landy, F. J., & Farr, J. L. (1980). Performance rating. *Psychological Bulletin, 87*, 72-107.

Niu, S. X., & Tienda, M. (2010). Minority student academic performance under the uniform admission law: Evidence from the university of texas at austin. *Educational Evaluation and Policy Analysis, 32*, 44-69. doi:10.3102/0162373709360063

Pearson, B. Z. (1993). Predictive validity of the scholastic aptitude test (SAT) for Hispanic bilingual students. *Hispanic Journal of Behavioral Sciences,15,* 342-356. doi:10.1177/07399863930153004

Pettigrew, T. F. (1979). The ultimate attribution error: Extending Allport's cognitive analysis of prejudice. *Personality and Social Psychology Bulletin, 5*, 461-476.

Pinel, E. C., Warner, L. R., & Chua, P. P. (2005). Getting there is only half the battle: Stigma consciousness and maintaining diversity in higher education. *Journal of Social Issues, 61*, 481-506.

Pulakos, E. D., White, L. A., Oppler, S. H., & Borman, W. C. (1989). Examination of race and sex effects on performance ratings. *Journal of Applied Psychology, 74*, 770-780.

Roth, P. L., Huffcutt, A. I., & Bobko, P. (2003). Ethnic group differences in measures of

job performance: A new meta-analysis. *Journal of Applied Psychology, 88,* 694-706. doi:10.1037/0021-9010.88.4.694

Rozin, P. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review, 5,* 296-320.

Ryan, A. M., McFarland, L., Baron, H., & Page, R. (1999). An international look at selection practices: Nation and culture as explanations for variability in practice. *Personnel Psychology, 52*, 359-391.

Sackett, P. R., Kuncel, N. R., Arneson, J. J., Cooper, S. R., & Waters, S. D. (2009). Does socioeconomic status explain the relationship between admissions tests and post-secondary academic performance? *Psychological Bulletin, 135,* 1-22. doi:10.1037/a0013978

Steele, C.M., & Aronson, J. (1995). Stereotype threat and the intellectual performance of African Americans. *Journal of Personality and Social Psychology, 69,* 797-811.

Stricker, L. J., Rock, D. A., Burton, N. W., Muraki, E., & Jirele, T. J. (1994). Adjusting college grade point average criteria for variations in grading standards: A comparison of methods. *Journal of Applied Psychology, 79,* 178-183.

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin, & S. Worchel (Eds.) *The Social Psychology of Intergroup Relations* (pp. 94-109), Monterey, CA: Brooks/Cole.

Taylor, M. C. (1979). Race, sex, and the expression of self-fulfilling prophesies in a laboratory teaching situation. *Journal of Personality and Social Psychology, 6,* 897-912.

Tenenbaum, H. R., & Ruck, M. D. (2007). Are teachers' expectations different for racial minority than for European American students? A meta-analysis. *Journal of Educational Psychology, 99*, 253-273.

Vaquera, G., & Maestas, R. (2008). Pre-college factors impacting persistence at a diverse university. *Journal of College Student Retention: Research, Theory and Practice, 10*, 425-445. doi:10.2190/CS.10.4.b

Vartanian, T. P., Karen, D., Buck, P. W., & Cadge, W. (2007). Early factors leading to college graduation for Asians and non-Asians in the United States. *The Sociological Quarterly, 48*, 165-165-197. doi:10.1111/j.1533-8525.2007.00075.x

Washington, V. (1982). Grade level differences in teacher perceptions of students by race and sex. *Journal of Applied Developmental Psychology, 3*, 81-83. doi:10.1016/0193-3973(82)90034-X

Willingham, W. W. (1985). *Success in college: The role of personal qualities and academic ability*. New York: College Entrance Examination Board.

Wing, J. Y. (2007). Beyond Black and White: The model minority myth and the invisibility of Asian American students. *The Urban Review, 39*, 455-487. doi:10.1007/s11256-007-0058-6

Young, A., Johnson, G., Hawthorne, M., & Pugh, J. (2011). Cultural predictors of academic motivation and achievement: A self-deterministic approach. *College Student Journal, 45*, 151-163.

Young, J. W. (1993). Grade adjustment methods. *Review of Educational Research, 63*, 151-165. doi: 10.2307/1170471

Young, J. W., & Kobrin, J. L. (2001). *Differential validity, differential prediction, and college admission testing: A comprehensive review and analysis* (College Board Research Report No. 2001-6). New York: College Board.

Yazedjian, A., Toews, M. L., & Navarro, A. (2009). Exploring parental factors, adjustment, and academic achievement among White and Hispanic college students. *Journal of College Student Development*, *50*, 458-467. doi:10.1353/csd.0.0080

Zwick, R., & Sklar, J. C. (2005). Predicting college grades and degree completion using high school grades and SAT scores: The role of student ethnicity and first language. *American Educational Research Journal*, *42*, 439-464. doi:10.3102/00028312042003439

APPENDIX A

DESCRIPTIVE STATISTICS AND CORRELATIONS BETWEEN ALTERNATIVE EXPLANATORY FACTORS

Table A-1

Descriptive Statistics and Correlations between Alternative Explanatory Factors for All Students

		M	*SD*	1	2	3	4	5
1	Course Load	33.30	15.30					
2	Cumulative College GPA	2.84	0.73	.427**				
3	English as Best Language	0.93	0.26	-.016**	-.002			
4	High School GPA	3.56	0.53	.157**	.412**	.013**		
5	SAT	1157.55	165.43	.139**	.384**	.053**	.434**	
6	SES	0.00	0.79	.058**	.151**	.215**	.121**	.279**

Note. $*p < .05$ (2-tailed)

Table A-2

Descriptive Statistics and Correlations between Alternative Explanatory Factors for Asian Students

		M	*SD*	1	2	3	4	5
1	Course Load	36.08	13.59					
2	Cumulative College GPA	2.96	0.65	.348**				
3	English as Best Language	0.74	0.44	-.004	.040**			
4	High School GPA	3.63	0.46	.151**	.302**	.030**		
5	SAT	1202.29	162.27	.120**	.356**	.200**	.288**	
6	SES	-0.20	0.88	.005	.072**	.232**	.109**	.301**

Note. $*p < .05$ (2-tailed)

Table A-3

Descriptive Statistics and Correlations between Alternative Explanatory Factors for Black Students

		M	*SD*	1	2	3	4	5
1	Course Load	32.70	15.02					
2	Cumulative College GPA	2.49	0.74	.494**				
3	English as Best Language	0.93	0.25	-.026*	-.116**			
4	High School GPA	3.36	0.56	.215**	.332**	.004		
5	SAT	1047.13	175.36	.207**	.419**	-.114**	.377**	
6	SES	-0.46	0.88	.093**	.028*	.235**	.090**	.162**

Note. $*p < .05$ (2-tailed)

Table A-4

Descriptive Statistics and Correlations between Alternative Explanatory Factors for Hispanic Students

		M	*SD*	1	2	3	4	5
1	Course Load	31.42	15.97					
2	Cumulative College GPA	2.63	0.76	.475**				
3	English as Best Language	0.75	0.43	.013	.042**			
4	High School GPA	3.54	0.52	.159**	.345**	.058**		
5	SAT	1102.91	161.96	.157**	.351**	.196**	.335**	
6	SES	-0.35	0.91	.024*	.130**	.171**	.040**	.299**

Note. $*p < .05$ (2-tailed)

Table A-5

Descriptive Statistics and Correlations between Alternative Explanatory Factors for White Students

		M	*SD*	1	2	3	4	5
1	Course Load	33.10	15.61					
2	Cumulative College GPA	2.86	0.73	.431**				
3	English as Best Language	0.98	0.13	.005	-.012**			
4	High School GPA	3.58	0.53	.154**	.445**	.004		
5	SAT	1162.48	159.64	.130**	.359**	.019**	.477**	
6	SES	0.12	0.68	.089**	.183**	.027**	.123**	.302**

Note. $*p < .05$ (2-tailed)