Differential Grading in North Carolina Public High Schools

April 20, 2012

Prepared by:
Sam Rauschenberg
Master of Public Policy Candidate
Sanford School of Public Policy
Duke University

Faculty Advisor:
Helen F. Ladd
Edgar T. Thompson Distinguished Professor of Public Policy; Professor of Economics
Sanford School of Public Policy
Duke University

Table of Contents

Executive Summary

Background

High school grades play an important role in college admissions. However, a student who receives an A in a course in one school may not have the same subject mastery as a student who receives an A in the same course in another school. As a result, some students enter college less prepared than their grades indicate while other students with the same content understanding may be denied admission. This phenomenon, known as differential grading, occurs when students are held to different grading standards in courses with the same curriculum and content (Godfrey 2011).

In North Carolina, end-of-course tests (EOC) in high school courses provide one way to measure differential grading by providing a standardized, curriculum-aligned estimate of student learning across the state. Each student's test score represents 25 percent of the final course grade. Currently, the Department of Public Instruction (DPI) offers tests in Algebra I, English I, and Biology. North Carolina formerly offered EOC tests in Geometry, Algebra II, Chemistry, Physics, Physical Science, Economics and Civics, and U.S. History, but the General Assembly eliminated these tests in 2011, citing cost savings and a desire to reduce the standardized testing burden. With the removal of EOC testing from these classes, course grades are the sole measure of student learning. To date, no research has examined the relationship between course grades and EOC test scores in North Carolina and whether this relationship varies across the state by student, teacher, school, and district characteristics.

Differential grading may be the result of various factors, including differences in teacher grading standards, district grading policies, student characteristics, teacher quality, and curriculum adherence. High school teachers often have significant latitude in determining their grade distribution (Camara 1998). Whether intentional or unintentional, teachers may assign a student's grade to reflect effort, persistence, a personal relationship, or a desire to increase a student's chances for college admission, rather than objective performance in the course. These forms of differential grading reflect varied grading standards or grading bias. However, grading variation can also occur due to grade discrimination, whereby teachers assign grades at least partially based upon stereotypes of a student's innate characteristics rather than on student performance on a certain assessment or course. Racial, gender, or other stereotypes of student performance in school or a certain subject influence teacher views of students, which may affect grading (Madon et al 1998). For example, teachers may believe that boys are better at math than girls or that minority students perform worse in school than white students (Reyna 2000; Hyde & Jaffree 1998). These stereotypes may cause teachers of a certain race or gender to issue grades that are systematically different for students of a certain race or gender.

Systematic differential grading also may exist between schools and districts. In North Carolina, districts have significant flexibility in setting grading policies. Further, schools that emphasize "teaching to the test" may have better test performance without affecting course grades.

In sum, differential grading likely exists in high schools. However, no research to date has examined the relationship between grades and EOC tests to identify differential grading patterns.

Analysis

Using three years of data on course grades and EOC test scores in North Carolina, I examine whether differential grading exists between districts. I then use regression analysis to identify patterns by student, teacher, school, and district characteristics in Algebra I, English I, Geometry, Biology, and Civics and Economics. While this analysis shows differential grading patterns, it does not differentiate between grade discrimination based upon stereotypes and differential grading due to systematic differences in effort or test-taking ability by race or gender.

Results

In all five subjects, districts with similar EOC test score averages have average course grades that vary by as much as a letter grade, or 0.6 standard deviations. This finding indicates that students in some districts may receive preferential treatment in college admissions relative to students in other districts with similar content mastery as measured by the EOC test.

Overall, I find that student characteristics predict differential grading to a greater extent than teacher, school, or district characteristics. In all five subjects, female students earn grades that are 1.25 to 2.7 points higher on a 100-point grading scale than male students, holding test scores and student, teacher, and school characteristics constant. This difference is equivalent to 0.12 to 0.27 standard deviations in course grades. Differential grading by race varies between subjects, and as a whole, race is less influential than other student characteristics.

Students classified as having limited English proficiency (LEP) earn grades that are 1.25 to 2.22 points higher than their peers, all else constant. Twelfth graders earn statistically significant higher grades than students in other grade levels. In contrast, low-income students earn 0.99 and 1.94 points lower grades than other students in all five subjects.

Overall, teacher characteristics are less influential than student characteristics. The only statistically significant result was in Algebra I where female teachers give grades that are 0.04 standard deviations lower than male teachers, all else constant. Class size has a small but positive relationship with course grades in all classes.

School-level characteristics, such as racial composition, the threat of missing Adequate Yearly Progress (AYP), and school size, do not have statistically significant impacts on grading. However, differential grading does exist by district location. Wake County and Charlotte-Mecklenburg Schools give lower grades relative to other districts, holding all else constant. Conversely, districts in the rural mountains give higher grades.

Policy Implications

This research is useful for educators and policymakers for several reasons. First, it shows that course grades alone inconsistently measure student learning across North Carolina. Thus, policymakers should think carefully about consequences of removing EOC tests from the curriculum. Students in districts that give systematically higher grades have an advantage in college admissions relative to students in other districts who have the same ability and content knowledge. Additionally, female students may be admitted to college preferentially over male students with similar content knowledge and ability as measured by EOC tests. Conversely, low-income students may be penalized in college admissions. This differential grading cannot be attributed directly to grade discrimination, varied grading standards, or systematic difference in student effort. However, educators and policymakers should consider how the current system favors students from certain districts or who have certain characteristics.

The results also provide an interesting comparison with a similar study in Georgia (Clark 2008). Georgia's HOPE Scholarship provides a college tuition scholarship to high school students with a 3.0 GPA. Thus, teachers may face pressure from students and parents to inflate grades. North Carolina does not have a similar scholarship program. Georgia students receiving the same letter grade as North Carolina students performed about 10 percentile points lower on their respective EOC tests in all five subjects. Thus, the HOPE Scholarship may encourage grade inflation, but, for reasons discussed in Section VII, this comparison provides only descriptive evidence of this effect.

I. Introduction

High school grades play an important role in college admissions. However, a student who receives an A in a course in one school may not have the same subject mastery as a student who receives an A in the same course in another school. Many factors lead to this differential grading, including differences in teacher grading standards, district grading policies, student characteristics, teacher quality, and curriculum adherence. To address this variability, end-of-course tests (EOC) in high school courses provide a standardized measure of student learning because the exams and curriculum are aligned with course standards. However, no research has examined the relationship between course grades and EOC test scores in North Carolina and whether the level of concordance between grades and test scores varies across the state by student, teacher, school, and district characteristics.

Using three years of statewide data on grades and EOC scores in Algebra I, English I, Geometry, Biology, and Civics and Economics, I first examine the relationship between course grades and EOC scores and whether course grades are consistent across North Carolina districts with similar average EOC scores. I then examine whether certain types of students are graded differently than other students with similar EOC scores and whether teacher and school characteristics contribute to differential grading.

Overall, I find that districts with similar EOC test score averages have average course grades that vary by as much as a letter grade, or 0.6 standard deviations, in all five subjects. In addition, student characteristics are stronger predictors of differential grading than teacher, school, or district characteristics. Female, Limited English Proficient (LEP), and 12th grade students earn statistically significant higher grades than other students in all five subjects, holding test scores and school characteristics constant. Low-income students, conversely, earn lower grades than other students in all five subjects.

This research is useful for educators and policymakers for several reasons. First, it provides insight into the consistency of course grades as a measure of student learning across the state. In addition, it demonstrates whether EOC tests provide an important source of information on learning beyond course grades. At the district level, the study presents evidence of significant grading variation in high schools between districts. At the student level, it shows that students with certain characteristics receive higher or lower grades despite having the same content mastery as measured by the EOC test. Students earning lower than predicted grades may be less likely to gain admission to college. Conversely, students who receive higher than predicted grades may need remedial work in college courses to relearn content knowledge, which has costly implications for the students and the state. Finally, the research provides a comparison with differential grading patterns in Georgia where the HOPE Scholarship's high school GPA requirement may increase the incentive for grade inflation relative to North Carolina.

Section II includes background information on EOC testing in North Carolina and a review of related literature. Section III describes the data used in this analysis. Section IV provides statewide and district-level descriptive statistics on the relationship between EOC test scores and course grades. Section V describes the empirical model. Section VI provides the regression results and discusses the findings. Section VII discusses policy implications from the results. Section VIII is a brief conclusion.

II. Background/Literature Review

End-of-Course Tests in North Carolina

First instituted in 1987, North Carolina EOC tests are aligned with the state's Standard Course of Study. North Carolina was one of the first states to adopt statewide, curriculum-based tests. As of the 2009-10 school year, 24 states had implemented or passed legislation to create some form of end-of-course testing (Zinth 2010).

In North Carolina, students on a full-year schedule take EOC tests in the last ten days of class, and those on a semester or block schedule take EOC tests in the last five days of class. Depending on the class, the exam has between 60 and 100 questions. Students receive three measures of their score: a scale score, a percentile rank, and an achievement level of I, II, III, or IV. The scale score is calculated from the raw score and, depending on the test, ranges from about 120 to 180. The percentile rank is calculated based upon how a student's score compares to students who took the test in the norming year rather than students in the same test administration. For each test in my analysis, the norming year occurred prior to the first year of data, so test scores are comparable across years. In terms of achievement levels, Level I indicates that a student lacks sufficient mastery of course standards, Level II indicates inconsistent mastery, Level III indicates consistent mastery, and Level IV demonstrates full mastery and advanced understanding. Prior to the 2010-11 year, the Department of Public Instruction (DPI) required students to earn at least Level III on the Algebra I, Biology, English I, U.S. History, and Civics and Economics tests to graduate. While DPI no longer requires a passing score on these EOC tests, some districts have chosen to retain it. In all cases, state law requires that a student's EOC score constitute 25 percent of the overall course grade. Each district sets up a formula to calculate a 100-point converted score from the scale score. This score appears on teacher score reports. Teachers are expected to enter this numeric score as 25 percent of each student's final grade.[1] DPI does not provide statewide recommendations or track the policies districts use to meet this requirement.

In recent years, the North Carolina General Assembly has reduced the number of EOC tests, citing cost savings and a desire to reduce the standardized testing burden. Prior to these

[1] To verify this practice, I spoke to central office officials and teachers in several North Carolina districts.

actions, the state offered EOC tests in Algebra I, Geometry, Algebra II, Biology, Chemistry, Physics, Physical Science, English I, Economics and Civics, and U.S. History. In 2009, it passed Senate Bill 202 to eliminate funding for the Physics and Chemistry tests. In 2011, the General Assembly passed House Bill 48 eliminating all standardized tests not required by federal law or as a condition of a federal grant. In high school, this legislation eliminated Algebra II, Civics and Economics, Physical Science, and U.S. History EOC tests, saving an estimated $2.73 million in the 2011-12 budget.[2] The State Board of Education voted to eliminate the Geometry EOC test in 2010, but its elimination is intended to be temporary as the state transitions to a new math curriculum. Due to these changes, the state is offering EOC tests in only Algebra I, Biology, and English I during the 2011-12 academic year at an estimated cost of $1.23 million. With the removal of EOC testing from the other classes, course grades have become the sole measure of student learning.

Literature Review

Grades play an important role in college admissions and merit scholarship eligibility. Many mid- to low-tier colleges make admissions decisions based almost exclusively on a student's GPA and SAT or ACT score. As a result, students have a strong incentive to improve their grades. While course grades are intended to measure a student's mastery of course content, other factors also influence a student's grade. For example, teachers in high schools often have significant latitude in determining their grade distribution. In a College Board survey of more than 15,000 high schools, 85 percent of high schools reported that teachers have significant grading flexibility (Camara 1998). Grading policies also differ across districts, and school administrations may have different expectations for the distribution of grades. As a result, a student's grade in a certain class in one school may not indicate the same mastery of course content as a student's identical grade in the

[2] Kristopher Nordstrom with the Fiscal Research Division of the North Carolina General Assembly Cost provided cost estimates.

same course in another school. This reality, often called differential grading, occurs when students are held to different grading standards in courses with the same curriculum and content (Godfrey 2011).

Whether intentional or unintentional, teachers may assign a student's score to reflect effort, persistence, a personal relationship, or a desire to increase a student's chances for college admission. If the course grade measures student persistence or hard work, then students with teachers of the same race or gender may be more motivated than students with teachers of a different race or gender. Furthermore, teachers may provide more support to students of the same race or gender, which may have a larger positive impact on their course grade than test score. Dee (2007) finds that students with same-gender teachers perform significantly better than students with opposite-gender teachers. Same-gender teachers also have better perceptions of same-gender students. Evans (1992) finds evidence that black students perform better on standardized tests of economic literacy when they have black teachers relative to teachers of another race. Similarly, at the community college level, Fairlie et al. (2011) find that underrepresented minority students have lower course withdrawal rates and higher pass rates when taught by a minority instructor.

Teachers also may change a student's grade based upon racial, gender, or other stereotypes of student performance in school or a certain subject. Madon, et al. (1998) find that a student's individual characteristics and past performance most influence teacher perceptions. However, gender, ethnic, and socioeconomics stereotypes do have some impact on how they view students. For example, teachers may believe that boys are better at math than girls or that minority students perform worse in school than white students (Reyna 2000; Hyde & Jaffree 1998). These stereotypes may cause grade discrimination, whereby a teacher assigns a grade at least partially based upon stereotypes of a student's innate characteristics rather than solely based upon student performance on a certain assessment or course.

Ehrenberg, et al. (1995) compare student learning gains over two years with teacher perceptions of their students' learning. They find that a teacher's race, gender, and ethnicity are more likely to influence their subjective evaluations of students than how much students actually learn as measured by a standardized test. Using surveys of teacher perspectives on each student, they find that teachers routinely rated female and white students higher than other students, holding test scores constant. Further, teachers of a certain race had more favorable views of ability for students of the same race relative to students of other races, holding test scores constant. Thus, the interaction of teacher and student characteristics may result in differential grading.

Most of the literature on differential grading at the student level has focused on differences by gender. One way researchers have identified the presence of differential grading is to compare how a student's teacher grades a test with how an identity-blind person grades the same or similar test. If the non-blind score systematically differs from the blind score, then teacher grading is susceptible to subjectivity. Lavy (2008) uses a difference-in-differences estimator to examine grading differences between genders on high school exams in nine courses in Israel. In each class, students took a teacher-graded school exam and an almost identical state exam within a few weeks that is graded blindly. Both scores are averaged to determine a student's overall matriculation score. Lavy finds that teacher-graded exam scores are higher than blind test scores for most tests in every subject area. More importantly, he finds a grading bias against boys of 0.05 to 0.25 standard deviations in the state exam score distribution across all subjects. These differences were sensitive to teacher characteristics, such as gender, age, experience, and family size, but were not sensitive to student characteristics, such as race, parent's education, or previous achievement. Thus, Lavy concludes that the grading difference can be attributed to an anti-male grading discrimination.

Hinnerich, et al. (2011) follow a similar methodology for a random sample of Swedish high school students' performance on a national Swedish language exam. The student's teacher and a blind grader from the Ministry of Education grade each test, so the study uses blind and non-blind grading of the same test rather than two tests. As a result, their results are not susceptible to differences in test structure or student behavior between test administrations, as is the case with Lavy's study. While they find that boys scored 15 percent lower than girls, they do not find evidence of discriminatory grading between genders. However, they find that blind grading yields scores that are 13 percent lower than non-blind grading for both genders, meaning that differential grading by classroom teachers does exist.

Another way to examine differential grading is to compare student performance on subject-based standardized tests with their teacher-assigned grades in those same subjects. Standardized tests provide a one-time snapshot of student's subject knowledge and test-taking skills while course grades reflect a more holistic picture of student performance as determined by multiple assessments, class participation, homework, attendance, and other factors. Grade comparisons between students with similar test scores can provide a picture of whether differential grading benefits certain students, assuming the tests are unbiased.

In Sweden, Lindahl (2007) compares student performance on national tests in Swedish, English, and Mathematics with teacher-assigned school leaving certificates, which are used for college admissions and job applications, to measure whether systematic differences exist by gender or native status. Both the tests and certificates evaluate students on the same scale of Fail (0), Pass (10), Pass with Distinction (15), and Pass Special Distinction (20). Teachers are supposed to use the test scores as a guide for school leaving certificates, but they have flexibility on a student-by-student basis. She finds that school leaving certificate averages are 0.02 to 0.06 standard deviations higher than test scores in all subjects. Using a difference-in-difference estimator with school and time effects, she finds that teachers assign higher grades to girls relative to boys in all subjects. Teachers also assign higher grades to non-native Swedish students relative to native students in mathematics and Swedish but not in English. The differences are largest in mathematics, making up 11 percent of a standard deviation in the grade different distribution for girls and 23 percent of a standard deviation for non-natives. Lindahl notes that these findings may only be generalizable to above average students because the data only included students with scores in all three subjects, eliminating lower performing students who may not have finished school. While grade discrimination may drive some of this difference, other factors, such as student effort and performance over the course of the school year, may have an impact on teacher grading.

Advanced Placement (AP) exams provide another way to compare course grades with a standardized measure of student mastery of course content. AP courses provide high-achieving high school students with college-level curriculum that culminates with a comprehensive exam for college credit. The exams are aligned with the curriculum in each subject.

Godfrey (2011) compares AP Exam scores and actual course grades in Biology, Calculus AB, English Literature, English Language, and U.S. History across five schools in each subject. She regresses AP exam scores on AP course grades with school fixed effects. She finds that the correlation between course grades and AP Exam scores varied widely within each subject and across schools. For example, in AP Biology, the correlation between exam score and grade varied from 0.29 to 0.77 across schools. Variation in course grades explained 27 percent of the variation in exam scores. In addition, a student's school explained from 19.4 to 32.5 percent of the variation in test scores. The other subjects demonstrated similar yet smaller patterns. AP Calculus AB had the smallest variation in the relationship between grades and test scores across schools, which may be due to the hierarchical and objective nature of mathematics relative to other subjects.

These results indicate that differences between schools may affect test scores, holding student grades constant. In other words, differential grading appears to exist across and within schools. However, as the author notes, the generalizability of results is limited because the data only include high school graduates or those who took AP exams. Thus, the evidence for differential grading cannot extend beyond high-achieving students, as is the case with Lindahl's study in Sweden.

Due to this limitation, an ideal standardized measure of student learning would be curriculum-based and required for all students. A statewide end-of-course (EOC) test fits this description. It is aligned directly with course curriculum to ensure that teachers focus class time on curriculum rather than also preparing students for other noncurriculum-based tests, such as a mathematics or English graduate exit exam. Finally, rather than a subsample of students, all students enrolled in a class with a corresponding EOC test must take the exam. In North Carolina, a student's EOC test score is 25 percent of the final course grade, which provides students with an immediate incentive to do well on it.

Several studies have compared EOC test scores with course grades. Troxell (2009) looks at the relationship between course grades and end-of-course tests in a Virginia school district in one year. She finds a linear relationship between grades and test scores but no differential grading.

In 1999, the Texas Education Agency examined the correlation between Algebra I course grades and end-of-course test scores from a representative sample in one year. Overall, 79 percent of students passed their course while only 45 percent of students passed the EOC test. By race, 48 percent of African American students and 41 percent of Hispanic students passed the course but not the EOC test. In contrast, only 30 percent of white students passed the course but not the EOC test. In terms of income level, 42 percent of low-income students passed the course but not the EOC test, which was 10 percentage points above other students. In sum, a larger proportion of minority and low-income students appear to be moving on to the next level of math without adequate preparation. The study also finds that course grades explain 35 percent of the variation in EOC test scores while ethnicity and income level explain only 6 percent of the variation.

While these studies provide some information on the relationship between course grades and EOC test scores, the analysis does not break down grading variation at the district, school, or teacher level to identify differential grading. Clark (2009) uses statewide data from Georgia to match course grades with Georgia EOC tests and finds sizable differential grading across districts in eight subjects. Georgia law requires the EOC test score to be 15 percent of a student's final course grade. Some school districts had a significant proportion of students receiving an A for a course even though they failed to meet standards on the EOC test. Other school districts had a large portion of students who received a C but exceeded standards on the exam. Humanities courses had larger differential grading than science and mathematics courses. This finding may be a result of humanities courses having a larger emphasis on writing, which is inherently more subjective to grade than math problems. Finally, each course had a higher EOC failure rate than course failure rate. The differences ranged from 8.18 percentage points in 11th grade English to 29.98 percentage points in Economics. Thus, each course had a significant number of students who passed the course but failed the EOC test, which mirrors the pattern found with Algebra I students in Texas.

This differential grading is particularly important in Georgia because students with a 3.0 high school GPA are eligible to receive a full tuition HOPE scholarship to a public university. As a result, differential grading creates unequal standards for scholarship eligibility and provides an inconsistent picture of how well students are prepared for college. Worse yet, it harms some students while favoring others. While this research points out the existence of grading subjectivity, it does not examine whether differential grading varies by school, teacher, or student characteristics.

In sum, the current literature indicates that differential grading likely exists in high schools. However, further research using course grades and mandatory, curriculum-based EOC tests in multiple years would provide a more complete picture of differential grading. Using three years of data on course grades and EOC test scores in North Carolina, I examine whether differential grading exists in North Carolina and identify patterns by student, teacher, and school characteristics. This paper adds to the debate in the literature on whether differential grading can be attributed to grade discrimination or other differences between students, teachers, and schools. The next section describes the data used for this analysis.

III. Data

I use statewide, student-level data for the 2007-08 to 2009-10 academic years from the North Carolina Education Research Data Center (NCERDC) at the Sanford School of Public Policy at Duke University. The data include student scores on five high school EOC tests (Algebra I, Geometry, Biology, English I, and Civics and Economics) and each student's grade in the course corresponding with the EOC. In addition, the data include student-level information on race, free or reduced price lunch eligibility (FRL), exceptionality status, and English language learning (ELL) status. The NCERDC's unique student identifier allows student EOC scores, course grades, and demographics to be linked between datasets and across the three years. Most districts reported courses grades on a numeric 1-100 scale or letter grades on a 7-point scale. However, several districts reported grades on a numeric 1-10 scale or a 7-point letter grade scale with plus and minus grades. I converted all grades to a numeric 1-100 scale to allow for comparison and inclusion as the dependent variable in my regression analysis.[3] Appendix 1 describes my methodology for converting grades to the same scale.

Using NCERDC's course membership data, I linked each student with the teacher in each course, enabling my regressions to include teacher race, gender, certification, and class size. However, depending on the class, only between 70 and 80 percent of the students had a matching teacher. Thus, I could only use this subset of data for any regression specifications that include

[3] Each regression specification includes an indicator variable to control for the grading scale.

teacher-level characteristics. Regressions without teacher characteristics include the entire high school student population with matching EOC test score and course grade. Appendix 3 includes summary statistics comparing student characteristics between the full student population and the subset. Finally, I supplemented the data with school- and district-level variables from NCERDC using the unique student identifier.

I restricted the data to include only North Carolina high school students. Since districts encourage students to take Algebra I in middle school, high schools in these districts will not have as many students in Algebra I as other courses. Some of these districts encourage only higher achieving students to take Algebra I in middle school. As a result, the statewide average EOC percentile score is lower in Algebra I than in other subjects. If teachers assign grades based upon a normal distribution, then high school students in these classes would be expected to receive higher course grades for a fixed test score than students in classrooms with the full distribution of students. Appendix 1 provides a more complete description of the methodology used to clean and merge the data.

IV. The Relationship between EOC Test Scores and Course Grades

Statewide Patterns

Differential grading occurs when students with similar knowledge and ability do not receive similar course grades. Before examining differential grading patterns, it is helpful first to understand the relationship between course grades and EOC test scores across the state. As a whole, grades and EOC scores should be highly correlated, especially because 25 percent of the course grade is the EOC score. Table 1 provides state-level summary statistics on this relationship over three years in the five courses examined in this paper. As noted above, the average Algebra I EOC score and course grade are lower than in other courses because students who take it in middle school, many of whom are higher achieving, are excluded from the data. For each course, a one standard deviation change in course grade is about 10 points on a 100-point scale.

Table 1: 2007-08 to 2009-10 Statewide EOC and Course Grade Summary Statistics

	Algebra I	English I	Geometry	Biology	Civics and Economics
Average EOC Percentile *Mean (SD)*	50.09 (25.31)	54.05 (27.13)	56.89 (26.82)	54.46(27.16)	57.83 (26.82)
Average Grade *Mean (SD)*	79.08 (10.03)	82.56 (9.58)	81.76 (9.75)	82.39 (9.50)	82.94 (9.78)
Correlation Coefficient	0.744	0.704	0.765	0.735	0.731
Number of Students	238,658	306,019	213,835	278,551	282,743
Number of Districts	115	115	115	115	115

*Calculations from NCERDC transcript and EOC data. Includes all non-alternative high school students with a matching course grade and EOC score. Percentile scores are calculated relative to the norming year, rather than other students in the same year.

Disaggregating the EOC scores by letter grade provides further information about the relationship between test scores and course grades. Table 2 provides the statewide average EOC percentile score and achievement level by letter grade. The letter grades are on a 7-point scale, where A is a 93-100, B is an 85-92, C is a 77-84, D is a 70-76, and F is below 70. Achievement Levels range from I to IV. Level I and II are considered failing while III and IV are considered passing. Overall, the average test scores in each letter grade are monotonic and consistent across courses. However, the percentile scores are much lower than the numeric grade required for a student to receive a certain letter grade. For example, the cutoff for a C is 77, but the percentile rank across subjects ranges from 47.4 to 52.3.

Table 2 also counts the percentage of students in each letter grade category who failed the EOC. For example, in Algebra I, 1 percent of students who received an A in the course failed the EOC test. Similarly, 90.5 percent of students who received an F in the course also failed the EOC test. Thus, nearly 10 percent of students passed the EOC test but received an F in the course. The percent of students in each letter grade category that failed the EOC is similar in every course except

English I. Roughly 7 to 10 percentage points fewer "D" and "F" students in English I failed the EOC test than in other courses. The English I EOC test includes composition and textual analysis sections, both of which test reading and writing ability rather than subject specific content. Other subjects require more content knowledge and skills, so students who do not learn coursework may be more penalized on those EOC tests than in English I.

Table 2: 2007-08 to 2009-10 EOC and Grade Statewide Summary Statistics by Letter Grade

	Algebra I	English I	Geometry	Biology	Civics and Economics
EOC Percentile Score for "A" Students	81.9	83.3	87.4	85.6	85.3
EOC Percentile Score for "B" Students	66.9	64.3	70.9	65.8	67.7
EOC Percentile Score for "C" Students	51.9	47.4	52.3	47.9	51.0
EOC Percentile Score for "D" Students	39.0	35.2	36.9	34.7	38.2
EOC Percentile Score for "F" Students	18.1	18.0	19.3	16.5	18.4
EOC Achievement Level for "A" Students*	3.8	3.8	3.9	3.8	3.7
EOC Achievement Level for "B" Students	3.3	3.4	3.5	3.3	3.2
EOC Achievement Level for "C" Students	2.9	2.9	2.9	2.8	2.8
EOC Achievement Level for "D" Students	2.4	2.6	2.5	2.5	2.4
EOC Achievement Level for "F" Students	1.6	1.9	1.8	1.8	1.6
Percent of "A" students who failed EOC**	1.0%	0.7%	0.4%	0.4%	0.8%
Percent of "B" students who failed EOC	6.8%	5.2%	4.2%	5.7%	6.7%
Percent of "C" students who failed EOC	24.8%	20.3%	21.9%	23.7%	27.0%
Percent of "D" students who failed EOC	51.3%	42.0%	51.0%	48.7%	52.2%
Percent of "F" students who failed EOC	90.5%	79.7%	87.2%	86.6%	89.0%
Percent of Students who failed EOC	34.7%	21.1%	25.4%	24.1%	25.2%
Percent of Students who failed course	15.6%	8.8%	9.6%	8.2%	8.6%
Percentage Point Gap	19.0	12.3	15.7	15.9	16.6

*Achievement Level ranges from 1 to 4.
**Students who failed EOC earned a Level I or II score.

The final part of Table 2 shows the gap between the EOC and course failure rates in each course. Relative to other subjects, Algebra I has a higher course and EOC failure rate because higher achieving students enroll in middle school. In addition, the gap between failure rates is larger than other subjects. While more than a third of Algebra I high school test takers failed the EOC, only 15.6 percent of them failed the course. This larger gap indicates that teachers may anchor grades around a certain distribution to some extent, regardless of ability or performance.

District-Level Patterns

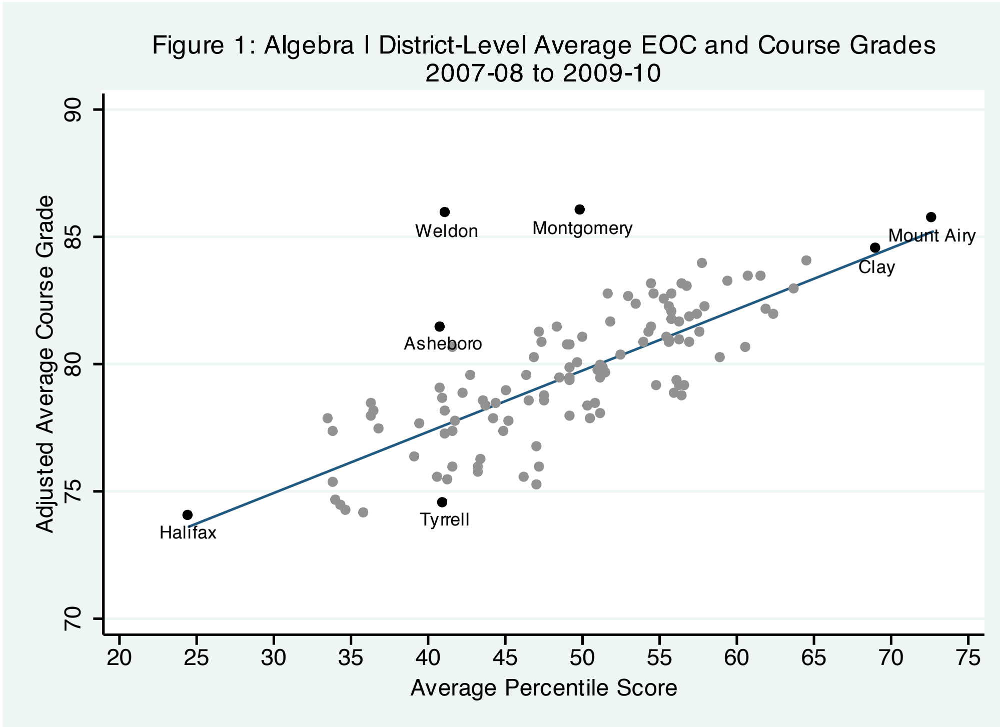
The summary statistics in the previous subsection provide a picture of the overall relationship between EOC scores and course grades, but district-level averages demonstrate whether this relationship holds for all students across the state. By aggregating the student-level data to the district level, I compare the adjusted average EOC score and course grade in each of the 115 North Carolina school districts using a scatterplot for each subject.[4] Figures 1 and 2 show the plots for Algebra I and English I. Appendix 2 includes scatterplots for Biology, Geometry, and Civics and Economics. Each class shows similar patterns to Algebra I and English I.

Each graph includes a data point representing the average course grade and average EOC percentile score in each of the 115 school districts. I use all three years of data to increase the number of students in each district and reduce variation that can occur with one year of data.[5]

---

[4] The regression analysis later in this study finds that the 7 districts that reported grades on a 1-10 numeric scale had grades that were 4.5 to 5 points lower than all other districts. Thus, I added the coefficient from the indicator variable for that grading scale from Model 1 of each subject's regression. This adjustment prevents those districts from appearing to give uncharacteristically lower grades than all other districts.

[5] Scatterplots of each year showed similar patterns but with more variation, and some districts that were outliers in one year were not outliers once the data were aggregated.

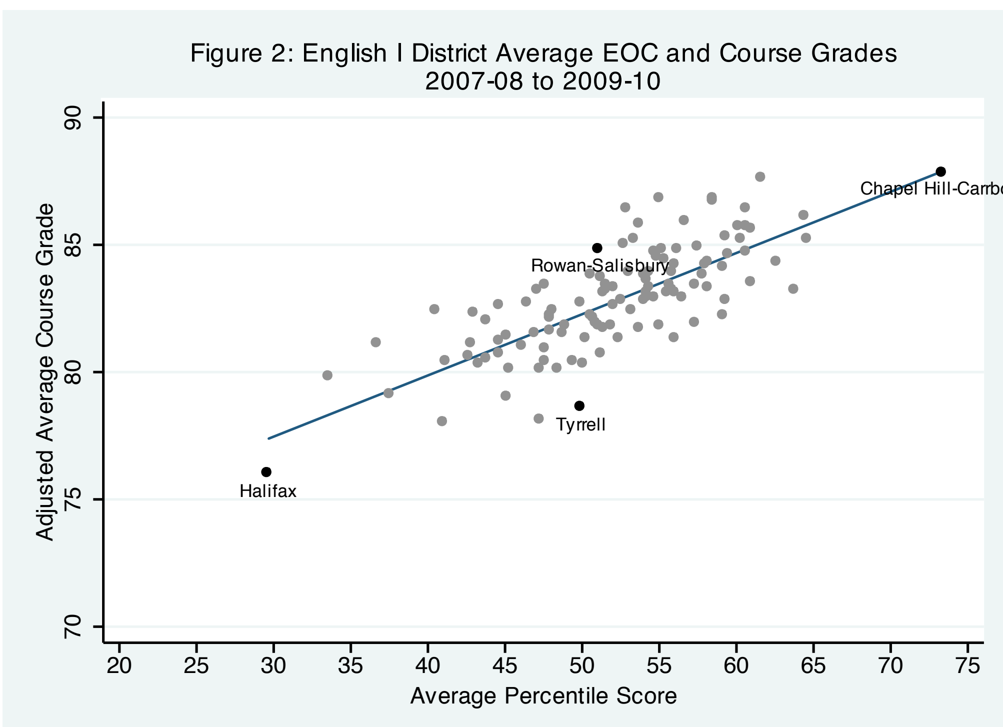


Both figures show considerable grading variation across districts with similar average EOC test scores. The line through the graph is the predicted regression line when course grades are regressed on EOC test scores. Districts above the line assign course grades that are higher than the predicted grade for their average EOC score. Districts below the line assign course grades that are lower than the predicted grade for their average EOC score.

As seen above, for a given scale score, the average district course grade for students in some districts varies by nearly a letter grade. For example, in Algebra I, Asheboro City Schools have an average percentile score of 41.0 and an average course grade of 81.4. At the same time, Tyrrell County Schools have an average percentile score of 41.1 and an average course grade of 74.5. Thus, students in Tyrrell Schools with the same level of achievement on the Algebra I EOC receive a grade 6.9 points below students in Asheboro City, nearly one letter grade. Put differently, an Algebra I student in Asheboro with an EOC percentile score of 41.0 receives a course grade of C while an Algebra I student in Rockingham with almost the same test score receives a D.

Several other patterns are worth mentioning. Weldon City and Montgomery County assign course grades more than a letter grade above the predicted value for their average EOC test score. When split into individual years, the higher grades in Weldon are driven by 2007-08 and 2008-09 because it is no longer an outlier in 2009-10. The pattern in Montgomery County is similar in all three years. Halifax County, Clay County, and Mount Airy Schools are outliers relative to other districts in terms of test scores, but their respective average grades are close to the predicted course grade.

The same pattern exists in English I. Rowan-Salisbury Schools has an average percentile score of 51.2 but an average course grade of 84.8. Tyrrell County Schools has an average percentile score of 50.0 but an average course grade of 78.6. Thus, the average student in Tyrrell County receives a grade 6.2 points lower than the average student in Rowan-Salisbury, despite having nearly identical EOCT test scores.



These findings are particularly interesting given that 25 percent of the course grade is the same scale score. Thus, the differences between district average grades arise either from the other 75 percent of a student's grade or differences in how districts convert the scale score into a numeric grade for teachers to input. Without the EOC test, the differential grading between districts may be even greater.

Multiple factors could be driving district-level grading variation across the state. Students in districts that emphasize "teaching to the test" may perform well on EOC tests without completing classwork or learning the material to the extent necessary to receive a course grade matching their test score. In this case, the test score provides an inflated perception of a student's grade, causing the district's average course grade to be lower than the predicted course grade for its test score. At the same time, some districts prohibit teachers from issuing grades below 60 on any assignment to allow failing students the opportunity to catch up. As a result, students may receive grades that are artificially higher than their ability or performance, causing the district's average grade to be higher than the predicted average course grade. Additionally, some districts may emphasize grading policies that put more weight on student work that is not directly related to test achievement, such as homework completion, attendance, or projects. Weighting grades heavily toward these factors would have an unclear impact on a student's grade depending on whether students complete the work.

Districts also have significant latitude in how they meet the state requirement that EOC scores be 25 percent of a student's final grade. Districts use the Department of Public Instruction's (DPI) WinScan software to convert the scale score into a 100-point grade, but DPI does not provide guidance or oversight for the formula used in this conversation. Further, some districts require teachers to fail students who receive a score that is more than a standard deviation below Level III while others allow students to pass the course as long as the final grade is above 70.

In sum, districts have significant latitude in determining grading policies, which may result in differential grading across the state. However, regression analysis is needed to determine if these differences between districts drive this differential grading or if other factors, such as student, teacher, and school characteristics, are more influential.

V. Model Specification

The descriptive statistics in the previous section establish that differential grading exists across North Carolina school districts. However, regression analysis with individual students as the unit of observation is necessary to identify grading patterns by student, teacher, school, and district characteristics, controlling for the student's test score. The NCERDC data allow for student level regression analysis controlling for these factors. To provide a baseline regression, Model (1) regresses a student's numeric course grade on his or her test score. The regression is specified as follows:

$$(1) \qquad G_{i,t} = \beta_0 + \beta_1 * T_{i,t} + \beta_2 * GS_d + \lambda_t + \varepsilon_{i,t}$$

where $G_{i,t}$ is student i's course numeric grade from 60 to 100 percent in year t, $T_{i,t}$ is a vector of four linear splines of student i's EOC scale score in year t, GS_d is a vector of three grading scale variables used to indicate the grading scale reported by district d, λ_t represents time effects, and $\varepsilon_{i,t}$ is the error term.

The linear splines in $T_{i,t}$ are split by the four achievement levels. Each spline equals the amount of the scale score that falls into its achievement level. For example, in Algebra I, the range for Level I is 139 or less, Level II is 140 to 147, Level III is 148 to 157, and Level IV is 158 or greater. A student with a score of 160 is in Level IV by 3 points. Therefore, his Level I spline equals

139, Level II spline equals 8, the Level III spline equals 10, and Level IV spline equals 3 points. A student with a score of 145 is in Level II by 6 points. As a result, her Level I spline is 139 and Level II spline is 6. Level III and Level IV are zero because his scores fall below these levels. This flexible functional form allows the slope of the regression line to vary within each achievement level. In other words, it assumes that the relationship between EOC scores and course grades does not remain constant across achievement levels. Thus, this functional form captures the effect of test scores on course grade more completely than a linear functional form. In each specification, the coefficient on each spline was statistically significant and different from the other spline coefficients.

I include the vector of GS variables in every model to control for how the grades are reported in the NCERDC. These grading scales are a numeric 1-10 scale, 7-point letter scale, and 7-point +/- letter scale. The omitted category is a numeric scale from 1-100. Appendix 1 discusses the distribution of grading scales by district and how I converted all grades to a numeric scale from 60 to 100. The coefficient on each grading scale variable shows the impact of a certain grading scale on a student's course grade relative to students on a numeric 1-100 scale.

The time effects, λ_t, hold constant all statewide differences across the three years of data. The 2007-08 year is the omitted category.

To begin controlling for student characteristics, Model (2) adds variables for student grade level, race, and gender to the regression.

$$(2) \qquad G_{i,t} = \beta_0 + \beta_1 * T_{i,t} + \beta_2 * GS_d + \delta_1 * GL_{i,t} + \delta_2 * female_i + \delta_3 * race_i + \delta_4 * (female_i * race_i) + \lambda_t + \varepsilon_{i,t}$$

where $GL_{i,t}$ is student i's grade level between 9th and 12th grade, $female_i$ is an indicator variable that equals 1 if student i is female and 0 otherwise, $race_i$ is a vector of indicator variables for student i's race, and $female_i * race_i$ is a vector of interaction terms between student i's race and gender.

Controlling for a student's grade level provides some indication of whether students are systematically graded differently based upon their grade level. I omit the year that students typically take the course across the state, which is 9^{th} grade for Algebra I and English I and 10^{th} grade for Geometry, Biology, and Civics and Economics. Since 12^{th} graders have added pressure to pass a course and schools have added incentive for them to graduate, I expect the coefficient on 12^{th} grade to be positive for each course.

The model includes three race indicator variables, Black, Hispanic, and Other. The "Other" variable includes American Indian and multiracial students. Minority students may earn lower grades for a fixed ability level if they believe the benefit to getting higher grades is lower than white or Asian students. Arcidiacono et al. (2010) estimate that high school grades are not an important labor market signal for students who do not attend college. Thus, students without college aspirations may have a lower incentive to earn course grades above what is required to pass a course. If fewer minority students have college aspirations than white students, then they may earn lower course grades than students with similar ability as measured by the EOC test.

I also include interaction terms between race and gender to measure whether male and female students within each race are graded differently. White and Asian male students are the omitted category.

Because the model interacts the three race variables with the female variable, the coefficient on female, δ_2, indicates any grading difference between white or Asian females and white or Asian males, holding test scores, year, race, and grading scale constant. If girls receive higher grades than boys, as found in the literature, then δ_2 will have a positive sign.

The coefficient on each race variable, δ_3, indicates any differential grading of male students by race, holding test scores, year, and grading scale constant. If δ_3 is negative for a certain category, then male students in this category receive lower grades than white male students. The coefficient on each interaction term is the marginal grading difference for being both female and a certain race. If girls receive higher grades than boys in each race, then these coefficients will have a positive sign.

Model (3) adds other student characteristics that may influence a student's grade, as follows:

$$(3) \qquad G_{i,t} = \beta_0 + \beta_1{*}T_{i,t} + \beta_2{*}GS_d + \delta_1{*}GL_{i,t} + \delta_2{*}female_i + \delta_3{*}race_i + \delta_4{*}(female_i{*}race_i) + \delta_5{*}frl_i + \delta_6{*}lep_i + \delta_7{*}except_i + \lambda_t + \varepsilon_{i,t}$$

where frl_i is an indicator variable for whether student i ever qualifies for free or reduced price lunch from 2006 to 2010.[6] The lep_i variable is an indicator for whether student i is classified as Limited English Proficient (LEP) in any of the three years of data. The $except_i$ variable is an indicator for whether student i has one of 16 exceptionalities in any of the three years of data. These exceptionalities include autism, behavior disorders, language or mental impairments, and physical disabilities.

FRL eligibility is correlated with race, so decoupling race and poverty indicates whether differential grading by race is driven by poverty. If the FRL variable is statistically significant and different from zero, it may reduce the magnitude and significance of race covariates. As with minority students, fewer low-income students may have college aspirations relative to other students. If so, the coefficient on this variable will be negative.

LEP eligibility is correlated with Hispanic students, so including this variable will show if the language barrier or other factors drive differential grading for Hispanics.[7] If the language barrier is most influential, then the coefficient on Hispanic from the previous specification will transfer to

[6] Since free or reduced price lunch (FRL) data is highly dependent upon student reporting, I used NCERDC data from 2006-2010 to create an indicator variable equal to 1 if a student was ever eligible for free or reduced price lunch over that time period. Students who are eligible for FRL in one year are likely to be similar to students eligible in all years. Thus, this variable decreases the volatility of FRL data due to selection, student movement, and other factors.

[7] While LEP and Hispanic are correlated, I used a VIF test for multicollinearity to determine whether both variables could be included in models. Both variables had VIF values below 2.5, so multicollinearity does not seem to be a problem.

LEP. Regardless of other characteristics, students with an exceptionality may receive higher grades than other students holding test score constant because teachers grade them more on effort than achievement. If so, the coefficient on $except_i$ will be positive.

In addition to student characteristics, grading patterns may exist by teacher and school characteristics. To measure this impact, Model (4) adds teacher- and school-level characteristics.

$$(4) \quad G_{i,t} = \beta_0 + \beta_1 * T_{i,t} + \beta_2 * GS_d + \delta_1 * GL_{i,t} + \delta_2 * female_i + \delta_3 * race_i + \delta_4 * (female_i * race_i) + \delta_5 * frl_i + \delta_6 * lep_i + \delta_7 * except_i$$
$$+ \gamma_1 * tchfemale_{i,j} + \gamma_2 * tchrace_{i,j} + \gamma_3 * teacher_{i,j} + \gamma_4 * clsize_{i,j} + \zeta_1 * school_{i,t} + + \zeta_2 * region_{i,t} + \lambda_t + \varepsilon_{i,t}$$

where $tchfemale_{i,j}$ is an indicator variable for whether student i's teacher j is female and $tchrace_{i,j}$ is a vector of three indicator variables for the race of student i's teacher j. These coefficients measure whether teachers of a certain race and gender grade students differently, holding constant test scores and student characteristics. $Teacher_{i,j}$ is a vector of teacher j's characteristics. These characteristics include indicator variables for whether teacher j has a temporary license, a Type 1 license, or a master's degree. Teachers with a Type 1 license have less than three years of experience. Teachers in each of these categories may grade students differently. For example, new teachers or those with a temporary license may be more susceptible to administrative pressure to alter grades or may seek to give the impression of student success to improve job prospects. $Clsize_{i,j}$ is the number of students in student i's class. As the class size increases, teachers may have less personal interaction with students and thus may be less susceptible to subjectivity. In addition, students may be less motivated to study due to decreased attention. If so, the coefficient on class size will be negative.

$School_{i,t}$ is a vector of school characteristics including percent of FRL students, percent of students who are black, Hispanic, or American Indian, school size, and an indicator variable for whether the school has missed AYP at least once over the three year period.[8] The racial and socioeconomics composition of the student body may affect teachers' racial stereotypes and thus how they grade minority students. The school's overall size may have an effect similar to that of class size in depersonalizing grading. In addition, teachers in large schools may feel less concentrated pressure from their administration to alter grades because several teachers teach the same course. If so, the coefficient on the school size will be negative. Finally, schools that are in threat of missing AYP may be more likely to encourage teachers to "teach to the test" since a large portion of AYP is student test scores. If this pressure occurs, then the coefficient would be negative on the AYP variable because students would receive higher test scores relative to their course grade. At the same time, principals may pressure teachers to inflate grades to increase graduation rates, which would have the opposite effect. Since test scores are a much larger portion of the AYP formula than graduation rates, I expect the coefficient on this variable to be negative.[9]

$Region_{i,t}$ is a vector of geographic indicator variables to measure whether grading patterns differ across the state by region. I split up districts using the same method as Clotfelter, et al. (2008). The groups are Rural Coast, Rural Mountains, Rural Piedmont, Urban Coast, Urban Mountains, and Urban Piedmont. Additionally, Charlotte-Mecklenburg, Cumberland, Guilford, Winston-Salem/Forsyth, and Wake County Schools are independent categories due to their large size relative to other school districts. Wake County is the omitted variable. The coefficient on each grouping will indicate if there are systematic grading differences between regions in the state.

Model 4 controls for school characteristics but does not take into account unmeasured differences between schools and districts that are constant over time. To address this shortcoming, Model (5) replaces school and district characteristics with school fixed effects.

---

[8] I use American Indian because the school-level data does not have an "other" category. It provides only the four racial categories listed above.

[9] Since schools that missed AYP and those that are close to missing it probably act in a similar manner, I created a one-year lag indicator variable for whether a school missed AYP in at least one year from 2006-07 to 2008-2009.

(5) $$G_{i,t} = \beta_0 + \beta_1 * T_{i,t} + \delta_1 * GL_{i,t} + \delta_2 * female_i + \delta_3 * race_i + \delta_4 * (female_i * race_i) + \delta_5 * frl_i + \delta_6 * lep_i + \delta_7 * except_i + \gamma_1 * tchfemale_{i,j} + \gamma_2 * tchrace_{i,j} + \gamma_3 * teacher_{i,j} + \gamma_4 * clsize_{i,j} + \phi_s + \lambda_t + \varepsilon_{i,t}$$

where ϕ_s is the school fixed effect. Thus, the remaining coefficients now measure differential grading within schools.

Differential grading may also occur between teachers and students of a certain race or gender. To measure whether teachers of a certain race or gender grade students of a certain race or gender differently, Model (6) adds interaction terms between teacher and student gender and race.

(6) $$G_{i,t} = \beta_0 + \beta_1 * T_{i,t} + \beta_2 * GS_d + \delta_1 * GL_{i,t} + \delta_2 * female_i + \delta_3 * race_i + \delta_5 * frl_i + \delta_6 * lep_i + \delta_7 * except_i + \gamma_1 * tchfemale_{i,j} + \gamma_1 * tchrace_{i,j} + \gamma_3 * teacher_{i,j} + \gamma_4 * clsize_{i,j} + \gamma_5 * (race_i * tchrace_{i,j}) + \gamma_6 * (female_i * tchfemale_{i,j}) + \zeta_1 * school_{i,t} + \zeta_2 * region_{i,t} + \lambda_t + \varepsilon_{i,t}$$

where $race_i * tchrace_{i,j}$ is a vector of nine interaction terms between the three student race and three teacher race variables. White or Asian students and teachers are the omitted category. $Female_i * tchfemale_{i,j}$ is the interaction term between the female student and female teacher interaction term. Male students and teachers are the omitted category. Model (6) also removes student race and gender interactions to reduce complexity. The rest of the model is the same as Model (4).

As with Model (4), the results in Model (6) may be susceptible to omitted variables at the school and district level. To address this issue, Model (7) uses school fixed effects instead of school and district characteristics.

(7) $$G_{i,t} = \beta_0 + \beta_1 * T_{i,t} + \delta_1 * GL_{i,t} + \delta_2 * female_i + \delta_3 * race_i + \delta_5 * frl_i + \delta_6 * lep_i + \delta_7 * except_i + \gamma_1 * tchfemale_{i,j} + \gamma_1 * tchrace_{i,j} + \gamma_3 * teacher_{i,j} + \gamma_4 * clsize_{i,j} + \gamma_5 * (race_i * tchrace_{i,j}) + \gamma_6 * (female_i * tchfemale_{i,j}) + \phi_s + \lambda_t + \varepsilon_{i,t}$$

where ϕ_s is the school fixed effect. Thus, the remaining coefficients measure differential grading within schools.

Each model includes robust standard errors that are clustered at the school level. The next section discusses results from each of these models. Appendix 3 provides descriptive statistics for the independent variables in each of the five courses.

VI. Regression Results

This section discusses the regression results for Algebra I and English I. Appendix 4 includes full regression tables for Algebra I, English I, Biology, Geometry, and Civics and Economics.[10] The following subsections are split to match the groups of variables in the regressions. Table 1 in Section IV provides the mean and standard deviation for course grades. In each course, the standard deviation for course grades was between 9.58 and 10.03. Thus, regardless of the course, a 1-point change in course grade is equivalent to about 0.1 standard deviations.

EOC Splines

Table 3 provides the coefficients on the EOC splines in Algebra I and English I in Model 1, the base regression model. The coefficients in all five subjects are statistically significant and fairly consistent across all models, even when school fixed effects are included in the model. The school effects capture unmeasured differences between schools, which include differences in how districts

[10] I also ran the following regressions with only the districts with the numeric 1-100 scale. The coefficients changed slightly, but the sign and significance were the same on every variable except Hispanic. In the subset regressions, Hispanic students earned grades that were about 0.3 points higher than in the full regressions. However, this difference represents only 0.03 standard deviations.

implement the 25 percent EOC requirement. Thus, the flexibility in policy implementation does not seem to have a significant impact on the relationship between EOC scores and course grades as expected.

The Achievement Level I and IV splines have smaller slopes than Level II or III in all subjects. A one-point increase in a student's EOC scale score relates to a 0.358-point grade increase for students in Achievement Level I and a 0.609-point increase for students in Achievement Level IV. The flatter Level I slope is probably because most students in this category also fail the course. Additionally, I changed all grades below 60 to equal 60, which also decreases the marginal effect of increases in EOC score within Level I. Level IV has a flatter slope, which is likely due to a ceiling effect for students at the top of the distribution.

Table 3: EOC Spline Regression Results

	Algebra I	English I
	Model 1	Model 1
	β / (SE)	β / (SE)
EOC Splines		
Ach Level I	0.358***	0.426***
	(0.02)	(0.02)
Ach Level II	1.242***	1.236***
	(0.02)	(0.02)
Ach Level III	0.929***	0.833***
	(0.01)	(0.01)
Ach Level IV	0.609***	0.594***
	(0.01)	(0.01)
R^2	0.574	0.519
N	239,578	305,878
F	2112.53	3377.14
P	0.000	0.000

* p<0.05, ** p<0.01, *** p<0.001
Dependent variable is a student's numeric course grade. Model 1 includes indicator variables for how districts reported grades and year effects. Standard errors are robust and clustered at the school level.

The slope of the Achievement Level II spline is 0.3 to 0.4 points steeper than the Achievement Level III spline in English I and Algebra I. For one-point scale score increase, an Achievement Level II student's grade increases by about 1.2 points in both subjects. This pattern may be due to teachers offering a grade bump to students who are on the bubble for passing the course. Most Level II scores are within one standard deviation of Level III, so some districts allow teachers to pass students in this category.

Student Characteristics

Differential grading also exists between student characteristics. Table 4 provides regression results in Algebra I for Models (2), (3), (4), and (5). Table 5 provides the same results for English I.

Table 4: Algebra I Student Regression Results

	Model 2	Model 3	Model 4	Model 5
	β / (SE)	β / (SE)	β / (SE)	β / (SE)
Student Characteristics				
10th grade	-0.906***	-0.928***	-0.989***	-0.723***
	(0.10)	(0.10)	(0.11)	(0.09)
11th grade	-0.494**	-0.554***	-0.793***	-0.331**
	(0.16)	(0.16)	(0.16)	(0.12)
12th grade	1.034***	0.928***	0.703**	1.153***
	(0.26)	(0.27)	(0.25)	(0.21)
Female	2.292***	2.352***	2.387***	2.406***
	(0.04)	(0.04)	(0.05)	(0.05)
Black	0.041	0.316*	0.696***	0.774***
	(0.14)	(0.14)	(0.07)	(0.07)
Hispanic	0.512***	-0.231	0.120	0.165
	(0.14)	(0.14)	(0.12)	(0.11)
Other Race	0.023	0.211	0.158	0.117
	(0.16)	(0.17)	(0.12)	(0.12)
Black Female Interaction	-0.548***	-0.541***	-0.481***	-0.493***
	(0.07)	(0.07)	(0.07)	(0.07)
Hispanic Female Interaction	-0.543***	-0.527***	-0.545***	-0.485***
	(0.11)	(0.11)	(0.12)	(0.11)
Other Race Female Interaction	-0.512**	-0.522**	-0.516**	-0.452*
	(0.16)	(0.16)	(0.18)	(0.18)
Free/Reduced Price Lunch		-0.586***	-1.052***	-1.111***
		(0.08)	(0.05)	(0.04)
Exceptionality		0.687***	0.761***	0.787***
		(0.09)	(0.09)	(0.08)
Limited English Proficient		2.092***	2.249***	2.214***
		(0.14)	(0.14)	(0.12)
R^2	0.586	0.589	0.609	0.64
N	239,578	239,578	186,711	186,711
F	1207.80	1210.12	877.12	1605.02
P	0.000	0.000	0.000	0.000

* p<0.05, ** p<0.01, *** p<0.001
Dependent variable is a student's numeric course grade.
Model 2 includes student grade level, race, and gender. It also includes year effects.
Model 3 adds other student characteristics.
Model 4 adds teacher, school and district characteristics.
Model 5 replaces school and district characteristics with school fixed effects.

White and Asian students are the omitted race category. Thus, coefficients represent the difference in grade in a specific group relative to this group.
Standard errors are robust and clustered at the school level.

Table 5: English 1 Student and Teacher Regression Results

	Model 2	Model 3	Model 4	Model 5
	β / (SE)	β / (SE)	β / (SE)	β / (SE)
Student Characteristics				
10th grade	0.858***	0.854***	0.300	0.407*
	(0.24)	(0.24)	(0.21)	(0.19)
11th grade	2.769***	2.603***	2.006***	2.005***
	(0.40)	(0.40)	(0.37)	(0.35)
12th grade	4.707***	4.539***	3.856***	3.542***
	(0.54)	(0.54)	(0.55)	(0.52)
Female	1.800***	1.849***	1.815***	1.825***
	(0.05)	(0.05)	(0.05)	(0.05)
Black	-1.227***	-0.649***	-0.366***	-0.320***
	(0.14)	(0.14)	(0.09)	(0.09)
Hispanic	-0.739***	-0.946***	-0.690***	-0.578***
	(0.14)	(0.16)	(0.14)	(0.13)
Other Race	-0.786**	-0.386	-0.712***	-0.659***
	(0.25)	(0.26)	(0.12)	(0.11)
Black Female Interaction	0.177*	0.183**	0.232**	0.202**
	(0.07)	(0.07)	(0.08)	(0.07)
Hispanic Female Interaction	0.357***	0.402***	0.421**	0.406**
	(0.11)	(0.11)	(0.13)	(0.12)
Other Race Female Interaction	0.007	0.005	0.096	0.084
	(0.13)	(0.13)	(0.14)	(0.14)
Free/Reduced Price Lunch		-1.514***	-1.904***	-1.947***
		(0.07)	(0.06)	(0.05)
Exceptionality		0.042	0.148	0.171
		(0.10)	(0.10)	(0.09)
Limited English Proficient		1.824***	2.099***	2.034***
		(0.15)	(0.16)	(0.14)
R^2	0.531	0.536	0.550	0.581
N	305,875	305,875	218,845	218,845
F	2111.67	1897.99	1015.00	1526.88
P	0.000	0.000	0.000	0.000

* p<0.05, ** p<0.01, *** p<0.001
Dependent variable is a student's numeric course grade.
Model 2 includes student grade level, race, and gender. It also includes year effects.
Model 3 adds other student characteristics.
Model 4 adds teacher, school, and district characteristics.
Model 5 replaces school and district characteristics with school fixed effects.

White and Asian students are the omitted race category. Thus, coefficients represent the difference in grade in a specific group relative to this group.

Standard errors are robust and clustered at the school level.

By grade level, Algebra I and English I have different patterns in grading. In Algebra I, 10^{th} graders receive grades 0.72 to 0.99 points lower than 9^{th} graders, holding all else constant. This difference may be due to the reality that most 10^{th} graders in Algebra I are either weaker students or are retaking the course. Those retaking the course may improve their test score from previous attempts, but if they do not change their habits in the classroom, their grade may not improve. Thus, they may receive a lower grade relative to 9^{th} graders, many of whom are taking it for the first time. Eleventh graders receive higher and statistically different grades than 10^{th} graders in every specification except Model 4. These values are still 0.33 to 0.79 points below 9^{th} graders. English I students, in contrast, receive higher grades in each subsequent grade level relative to 9^{th} graders. Students in 11^{th} grade earn a course grade that is 3.5 to 4.7 points, more than half a letter grade, higher than 9^{th} graders in English I.

Twelfth graders earn statistically significant higher grades than other grade levels in both English I and Algebra I. Some of this increase is likely due to increased student effort to graduate because they are closer to the end of school. Additionally, principals may pressure teachers to raise the grades of students who are in danger of failing so that they can graduate.

Differential grading also exists by gender and race. In all subjects, female students earn 1.25 to 2.7 points higher grades than male students regardless of race, holding test score constant. These results are not sensitive to other covariates or school fixed effects. This finding aligns with the literature on girls earning higher grades than boys. Girls may be more conscientious students by turning in more work, having fewer discipline problems, and studying more than boys. In addition, girls may respond more negatively to pressure on testing day, causing them to receive scores below their ability. Finally, teachers may have gender stereotypes that systematically cause them to give higher grades to girls.

In terms of race, opposite findings emerge for Algebra I and English I. All minority male students in Algebra I have positive coefficients relative to white male students when teacher and school characteristics are included in the model, but only the coefficient for black students is statistically significant. Black male students earn a grade 0.774 points higher than white or Asian male students, holding test score, teacher characteristics, and school constant. Furthermore, female black students earn a grade that is 2.687 points higher than white male or Asian students, a difference of 0.27 standard deviations. The black coefficient only becomes statistically significant when other student characteristics, including poverty, and school characteristics are added. Hispanic male students receive statistically significant higher grades only when the control for Limited English Proficient (LEP) students is not in the regression (Model 2). Since many Hispanic students may also be classified as LEP, it seems that their higher grades have more to do with their language classification than with their ethnicity.

In English I, the signs on all minority variables are negative and statistically significant. However, as with Algebra I, when teacher and school characteristics are added, the coefficient on the Black and Hispanic variable moves in a positive direction.

Each of the other student characteristics is also statistically significant and has the same sign in Algebra I and English I. FRL students receive grades that are 1.111 points lower than other students in Algebra I and 1.947 points lower in English I. Relative to other students, fewer low-income students probably have college aspirations. Since high school grades have a limited impact on the blue-collar labor market, they may not try to earn higher grades on the margin as long as they pass the course (Arcidiacono 2010).

Students with an exceptionality receive higher grades in both courses, but the effect is larger and statistically significant only in Algebra I. Teachers may try to compensate for academic exceptionalities by artificially improving grades. Or, for students with exceptionalities related to discipline, teachers may increase their grades so that they do not have to teach them again.

LEP students exhibit the largest positive sign in both classes. LEP students earn grades that are more than 2 points higher than their peers, all else constant. LEP students may benefit from the extra supports afforded to them in several ways. They have more one-on-one attention, which may increase the amount of classwork they complete. They also may have help on their classwork that allows them to earn higher grades than students with the same test score who do not have the same level of help. Lastly, teachers may try to compensate for the language barrier by artificially raising a student's grade.

The patterns for FRL, LEP, and students with an exceptionality hold true in Biology, Civics and Economics, and Geometry. These regression tables are included in Appendix 4.

In sum, across all subjects, the positive coefficients for female students, LEP students, and 12th graders indicate that they all receive higher grades for a fixed test score. The negative coefficient for low-income students indicate they receive lower grades for a given test score. The sign and significance of the race indicators varies across subjects, and the effect sizes are smaller than other student characteristics.

Teacher Characteristics

The regression results from Models (4) and (5) in Table 6 demonstrate that teacher characteristics as a whole do not play as large a role in differential grading as student characteristics. In Algebra I, female teachers give grades that are 0.383 points lower than male teachers, holding test scores, student characteristics, and school constant. However, this difference is only 0.04 standard deviations. Other teacher characteristics are not statistically significant. None of the teacher characteristics are statistically significant in English I.

In both classes, the class size variable has a small but statistically significant negative coefficient. For each additional student, a student's course grade decreases by 0.037 points in Algebra I and 0.045 points in English I, holding student, teacher, school, and district characteristics constant. As the class gets larger, students receive less personal attention from the teacher. Consequently, students may be less motivated to turn in class work, and teachers may be less likely to increase student grades because of their personal relationship.

Table 6: Teacher Regression Results

	Algebra I		English I	
	Model 4	Model 5	Model 4	Model 5
	β / (SE)	β / (SE)	β / (SE)	β / (SE)
Teacher/Class Characteristics				
Female	-0.528***	-0.383**	-0.129	0.058
	(0.14)	(0.13)	(0.18)	(0.17)
Black	0.078	0.017	0.066	0.045
	(0.20)	(0.20)	(0.24)	(0.22)
Hispanic	0.248	0.351	0.238	0.225
	(0.66)	(0.59)	(0.58)	(0.54)
Other Race	-0.481	-0.172	0.056	0.487
	(0.51)	(0.50)	(0.74)	(0.76)
Temporary License	0.308	0.179	0.443*	0.227
	(0.22)	(0.18)	(0.22)	(0.23)
Type 1 License	0.411*	0.17	0.077	0.226
	(0.17)	(0.17)	(0.17)	(0.15)
Masters	0.041	-0.071	-0.096	-0.02
	(0.13)	(0.13)	(0.14)	(0.13)
Class size	-0.043***	-0.045***	-0.029**	-0.037***
	(0.01)	(0.01)	(0.01)	(0.01)
R^2	0.609	0.64	0.55	0.581
N	186,711	186,711	218,845	218,845
F	877.12	1605.02	1015.00	1526.88
P	0.000	0.000	0.000	0.000

* p<0.05, ** p<0.01, *** p<0.001
Dependent variable is a student's numeric course grade.
Model 4 includes student, school, and district characteristics.
Model 5 replaces school and district characteristics with school fixed effects. Both models include year effects.

White and Asian teachers are the omitted race category. Thus, coefficients represent the difference in grade in a specific group relative to this group.

Standard errors are robust and clustered at the school level.

Teacher and Student Interactions

To fully capture differential grading at the student and teacher level, Models (6) and (7) include interaction terms between teacher and student race and gender. For ease of interpretation, Table 7 includes the calculated coefficients from the regression results with white and Asian male students and white and Asian male teachers as the omitted category. The coefficients are starred if the interaction term is statistically significant. Full regression results are included in Appendix 4.

Table 7: Differential Grading Calculations for Race and Gender (Based upon regression results in Appendix 4)

	Algebra I		English I	
	Model 6	Model 7	Model 6	Model 7
Race Interactions[1]				
Black student/Black teacher	0.617*	0.584*	0.047	0.033
Black student/Hispanic teacher	0.830	0.830	-0.083	-0.532
Black student/Other teacher	0.467	0.717	-0.029	0.188
Black student/White teacher	0.404	0.513	-0.362	-0.308
Hispanic student/Black teacher	-0.323	-0.325	-0.619	-0.436
Hispanic student/Hispanic teacher	-0.349	0.298	0.585	0.898
Hispanic student/Other teacher	-1.667	-1.081	0.109	0.835*
Hispanic student/White teacher	-0.106	-0.019	-0.5	-0.406
Other student/Black teacher	0.065	0.297	-0.362	-0.521
Other student/Hispanic teacher	0.699	0.560	-1.848	-2.439
Other student/Other teacher	-1.497	-0.827	-1.323	-0.707
Other student/White teacher	-0.087	-0.163	-0.715	-0.63
White student/Black teacher	-0.077	-0.028	-0.284	-0.248
White student/Hispanic teacher	0.206	0.353	0.108	0.391
White student/Other teacher	-0.343	0.053	0.285	0.877
White student/White teacher	Reference	Reference	Reference	Reference
Gender Interactions[2]				
Female student/Female teacher	1.586**	1.756**	1.654***	1.846***
Female student/Male teacher	1.997***	2.023***	1.573***	1.583***
Male student/Female teacher	-0.633***	-0.485**	-0.343	-0.148
Male student/Male teacher	Reference	Reference	Reference	Reference

* p<0.05, ** p<0.01, *** p<0.001
Dependent variable is a student's numeric course grade.
[1]Omitted Category in race interactions is white teacher/white student.
[2]Omitted category in Gender interactions is male teacher/male student.

Model 6 includes student, school, and district characteristics and year effects.
Model 7 replaces school and district characteristics with school fixed effects.

Standard errors are robust and clustered at the school level.

In Algebra I, black students with black teachers earn a grade that is 0.584 points higher than white students with white teachers, holding all else in the equation constant. This difference was statistically significant with p-value less than 0.05. The only other subject with a statistically significant positive coefficient is Geometry. Since this differential grading only exists in math courses, it does not appear that black teachers systematically grade students differently based upon their race.

In English I, the only race interaction with a statistically significant coefficient is for Hispanic students with teachers of "other" race. Since less than 1 percent of teachers are classified as "other," this finding is likely due to small sample size.

The interaction of teacher and student gender is more influential than race interactions. For female students in Algebra I, female teachers assign grades that are 0.26 points lower than male teachers, which was narrowly statistically different with p-value = 0.05. In English I, however, female teachers assign grades to female students that are 0.26 higher than male teachers but are not statistically different. Both results are statistically significant. For male students, female Algebra I teachers issue grades that are 0.485 points lower than male teachers. In English I, female teachers issue grades that are 0.148 points lower than male teachers. Only the Algebra I coefficient is statistically significant.

Since the results do not indicate patterns in all subjects, differential grading does not appear to occur systematically between teachers and students of a certain race and gender.

School Characteristics

Overall, school characteristics do not seem to play as important a role in differential grading as student characteristics. Results for Model 6 are shown in Table 8. For English I, the coefficient on the percent poverty variable is positive but small and statistically significant. For a 10-percentage point increase in a school's low-income students, a student's course grade increases by 0.2 points. The percent of students who are American Indian has a statistically significant coefficient of 0.059. It means that for every 10-percentage point increase in American Indian students, student grades increase by 0.59 points while holding individual characteristics and test scores constant. In Algebra I, the coefficient on the percent Indian variable is also positive and statistically significant, but it is smaller than in English I. The coefficient on school size is also statistically significant but small. As school size increases by 100 students, a student's grade decreases by 0.1 points. While this effect is small, it is in the expected direction because larger schools may be more impersonal than smaller schools.

Table 8: School Characteristics Regression Results

	Algebra I	English I
	Model 6	Model 6
	β / (SE)	β / (SE)
School Characteristics		
Percent Poverty	0.009	0.020*
	(0.01)	(0.01)
Percent Black	-0.004	-0.004
	(0.01)	(0.01)
Percent Hispanic	-0.005	-0.009
	(0.02)	(0.02)
Percent Indian	0.034**	0.059***
	(0.01)	(0.01)
Threat of missing AYP	0.182	-0.255
	(0.33)	(0.27)
School size (students)	-0.001***	0.000
	(0.00)	(0.00)

* p<0.05, ** p<0.01, *** p<0.001

Dependent variable is a student's numeric course grade.
Model 6 includes student, school, and district characteristics and year effects. Model 7 replaces school and district characteristics with school fixed effects.

District Characteristics

Differential grading also exists across geographic regions in the state, as seen in Table 9. Relative to Wake County Schools, all groups except Charlotte-Mecklenburg have positive signs on their coefficients. Guilford, Cumberland, and Winston-Salem/Forsyth issue grades between 1.358 and 4.197 points higher than Wake County in Algebra I and English I. Of the six geographic groupings, Rural Mountains is the only group with a positive coefficient that is statistically different from the other groupings in English I. Rural Mountains, Rural Piedmont, and Urban Mountains are larger and statistically different from the other groups in Algebra I. Course grades in Rural Mountain schools are higher than the other five geographic groups in every subject. Thus, holding test scores and student, teacher, and school characteristics constant, students in the rural mountains earn higher grades than students in the other geographic groups.

Each regression also includes three indicator variables to control for the method that districts used to report grades to the DPI. Districts on either letter grade system do not have statistically significant grading differences relative to the omitted category, a standard 1-100 numeric grading scale, holding constant test scores, student, teacher, school, and district characteristics. However, the seven districts that reported grades on the 1-10 numeric scale issued grades that are more than 4.5 points lower in all five subjects. This large difference is likely due to the inability to convert grades to a 1-100 scale with precision. For the purposes of this study, a student who received an 8 in a course in this district was converted to receive an 80. However, an 8 could represent a broad range of grades around 80. Because of the potential for error in conversion, these indicator variables are included in every regression model to hold the grade reporting method constant.[11] Appendix 1 provides a list of district grade reporting methods.

[11] As discussed earlier in the paper, regressions with only districts on the 1-100 numeric scale had similar results regressions with all districts.

Table 9: District Characteristics Regression Results

	Algebra I	English I
	Model 6	Model 6
	β / (SE)	β / (SE)
District Characteristics		
Charlotte-Mecklenburg	0.061	-1.183*
	(0.46)	(0.50)
Cumberland	1.691*	1.684*
	(0.77)	(0.73)
Guilford	4.197***	1.762**
	(0.55)	(0.67)
Winston-Salem/Forsyth	2.635***	1.537*
	(0.60)	(0.66)
Rural Coast	1.358*	0.575
	(0.60)	(0.62)
Rural Mountains	1.897***	1.459*
	(0.55)	(0.63)
Rural Piedmont	1.959***	0.894
	(0.51)	(0.55)
Urban Coast	1.131*	0.705
	(0.55)	(0.56)
Urban Mountains	2.039**	0.49
	(0.63)	(0.61)
Urban Piedmont	1.210*	0.768
	(0.60)	(0.54)
Numeric 1-10 Grade Scale	-4.496***	-4.951***
	(0.34)	(0.30)
7-Point Letter Grade (A-F)	0.234	0.078
	(0.32)	(0.32)
7-Point Letter Grade (+/- A-F)	0.237	-0.235
	(0.44)	(0.41)

* p<0.05, ** p<0.01, *** p<0.001
Dependent variable is a student's numeric course grade.
Model 6 includes student, school, and district characteristics and year effects. Model 7 replaces school and district characteristics with school fixed effects. What is the reference district?

Summary

Overall, student characteristics are stronger predictors of differential grading than teacher, school, or district characteristics. Female, Limited English Proficient, and 12th grade students earn statistically significant higher grades than other students in all subjects, holding test scores, student characteristics, teacher, and school characteristics constant. Low-income students, in contrast, earn lower grades than other students in all five subjects. Student race and teacher characteristics did not have a pattern across all subjects. If teachers alter student grades based upon student effort, discipline, or attendance, then these coefficients may be more indicative of these factors than systematic differential grading.

VII. Policy Implications

North Carolina EOC tests provide a unique opportunity to measure differential grading patterns across the state by comparing course grades and EOC test scores. As a result, this research has several important implications. First, it provides insight into whether course grades provide a consistent measure of student learning across the state. Descriptive statistics show that districts with similar test score averages have average course grades that vary by as much as a letter grade, or 0.6 standard deviations, across all subjects. Students in districts that give systematically higher grades have an advantage in college admissions relative to students in other districts who have the same ability and content knowledge. Furthermore, this grading variation also affects certain groups of students differently, even when controlling for EOC test performance. This differential grading can result in certain types of students earning higher or lower grades relative to other students with similar ability and knowledge as measured by test scores. Students earning lower than predicted grades may be penalized in college admissions. At the same time, students with higher than predicted grades may need additional remedial work to relearn material in college, which has costly implications for the students and the state. Consequently, EOC tests play an important role in measuring student mastery of course content consistently across the state. Removing EOC tests, as North Carolina has done in all courses but Algebra I, English I, and Biology, reduces the state's ability to compare student learning across the state and may exacerbate differential grading.

Finally, the results provide an interesting comparison with a similar study based on one year of data in Georgia (Clark 2008). Georgia students also take EOC tests in high school courses, but the scores make up only 15 percent of a student's final grade. Georgia's HOPE Scholarship provides a college tuition scholarship to high school students with a 3.0 GPA. Thus, teachers may face pressure from students and parents to inflate grades. North Carolina does not have a similar scholarship program. If this incentive has an impact on grades in Georgia, then the EOC percentile rank for students in each letter grade category should be lower than North Carolina students, particularly with "B" or "C" students because they are on the bubble for scholarship eligibility. In addition, gaps between EOC failure rates and course failure rates should be larger. Both patterns hold true.

Across the five subjects I examined, the average percentile rank for "B" students was 56.7 in Georgia and 67.1 in North Carolina. For "C" students, the percentile rank was 39.0 in Georgia and 50.1 in North Carolina.[12] Thus, students receiving similar course grades in Georgia performed about 10 percentile points lower than North Carolina students on their respective EOC tests. The gap between failure rates also was larger in Georgia in all subjects except Algebra I, ranging from a difference of 4.5 percentage points in English I to 13.4 percentage points in Civics and Economics.

Other factors beyond the HOPE Scholarship may influence the relationship between EOC scores and course grades. Since the EOC score makes up a smaller proportion of a student's final grade than North Carolina's 25 percent, the average percentile rank for students should be slightly

[12] These figures for North Carolina are included in Table 2 in Section IV.

lower. Statistically, the Georgia data also include only one year of observations while my North Carolina data is over three years. North Carolina students are ranked relative to the norming year rather than to students in the same testing year while Georgia students were ranked relative to students in the same testing year. Finally, other differences between states, such as EOC test structure and differences in EOC graduation requirements, also reduce the comparability across states. As a result of these issues, this comparison only provides descriptive evidence that the HOPE Scholarship creates an additional incentive for teachers to inflate student grades.

VIII. Conclusion

This research adds to the literature on differential grading by comparing mandatory EOC test scores with course grades for all high school students in North Carolina over three years. I find that course grades have a significant amount of variability between districts across the state. In addition, this variation differs by student characteristics, such as gender, grade level, family income status, and English language learning classification. One limitation of the research is that it cannot distinguish between causes of differential grading, such as varying grading standards, grade discrimination, and systematic differences in student behavior. Nonetheless, college admissions decisions often rest heavily on a student's GPA. The existence of differential grading benefits some students at the expense of other students with similar ability and content knowledge as measured by test scores.

Looking forward, EOC tests play an important role in identifying grading patterns and may mitigate some grading variation. Thus, educators and policymakers should consider the implications of removing them from the curriculum. Additionally, they should seek other ways to reduce differential grading and its impact on college admissions.

APPENDIX 1: Data Cleaning Methodology

This section discusses the methodology and assumptions I used to create the dataset for this analysis.

Student grades are not uniformly reported across all schools and districts. The North Carolina Department of Public Instruction (DPI) mandates that districts use a seven-point system, whereby an A is a 93-100, B is an 85-92, C is a 77-84, D is a 70-76, and F is any grade below a 70. Beyond this policy, districts have the freedom to report grades as numbers or letters. They can also split letter grades into a plus/minus scale. Table A1 provides the grading scales by district as reported in the NCERDC data. In 2008, 48 percent of the students received a numeric grade from 0 to 100, and 52 percent received a letter grade. Of students with a letter grade, less than 10 percent received a grade on the plus/minus scale.

To perform regression analysis with course grades as the dependent variable, I converted all grades to numeric grades on a 1-100 scale. I dropped all grades that did not correspond with a numeric or letter grade indicating that the student completed the course, such as P, W, WF, or S, were. In addition, I changed all numeric grades below 60 to equal 60, making the grade range from 60-100. Without this adjustment, students with failing grades far below 70 would bias estimates downward. In addition, differential grading is not important between failing grades because students do not receive credit for a course in both cases.

Table A1: North Carolina Grading Scales by District*

Numeric 1-10 Scale	Standard 7-Point Letter Scale	Plus/Minus 7-Point Letter Scale
Davidson	Alamance-Burlington	Ashe
Granville	Buncombe	Cabarrus**
Iredell-Statesville	Burke	Cleveland
Mooresville	Caldwell	Edgecombe
Orange	Cherokee	Polk
Rockingham	Forsyth	Robeson
Yadkin	Guilford	
	Haywood	
	Henderson	
	Charlotte-Mecklenburg	
	Chapel Hill-Carrboro	
	Person	
	Randolph	
	Asheboro	
	Rowan-Salisbury	
	Rutherford	
	Elkin	
	Vance	
	Wake	
	Wilkes	
	Wilson	

*All districts not listed used a 1-100 numeric scale. Grading scales were determined based upon the distribution of grades in the NCERDC data.

**Cabarrus switched to 1-100 numeric scale for 2009-10.

To convert letter grades to numeric grades, I imputed the average numeric grade in each letter grade category from all districts on the numeric scale. The average racial and socioeconomic composition of districts on a letter grade system and those on a numeric grade system were not statistically different. Assigning grades based upon the midpoint of the grade range for each letter would have assumed that actual numeric grades within each letter are distributed randomly. While this pattern holds true for B, C, and D, it does not hold true for A's and F's. The imputed value is about 1 point below the midpoint of the A range, likely because of a ceiling effect at 100. For F, the distribution is weighted heavily toward 60 because all grades below 60 are converted to 60. As a result, the imputed value is 62.49 instead of 65. For schools on the plus/minus letter grade system, I imputed the midpoint for the range with the exception of A+ because the range was only 2 to 3 points in each category. Due to the ceiling effect, I imputed 99 for A+ since the range was 99-100. The following table shows the ranges and imputed values.

Table A2: Conversions from Letter to Numeric Grades

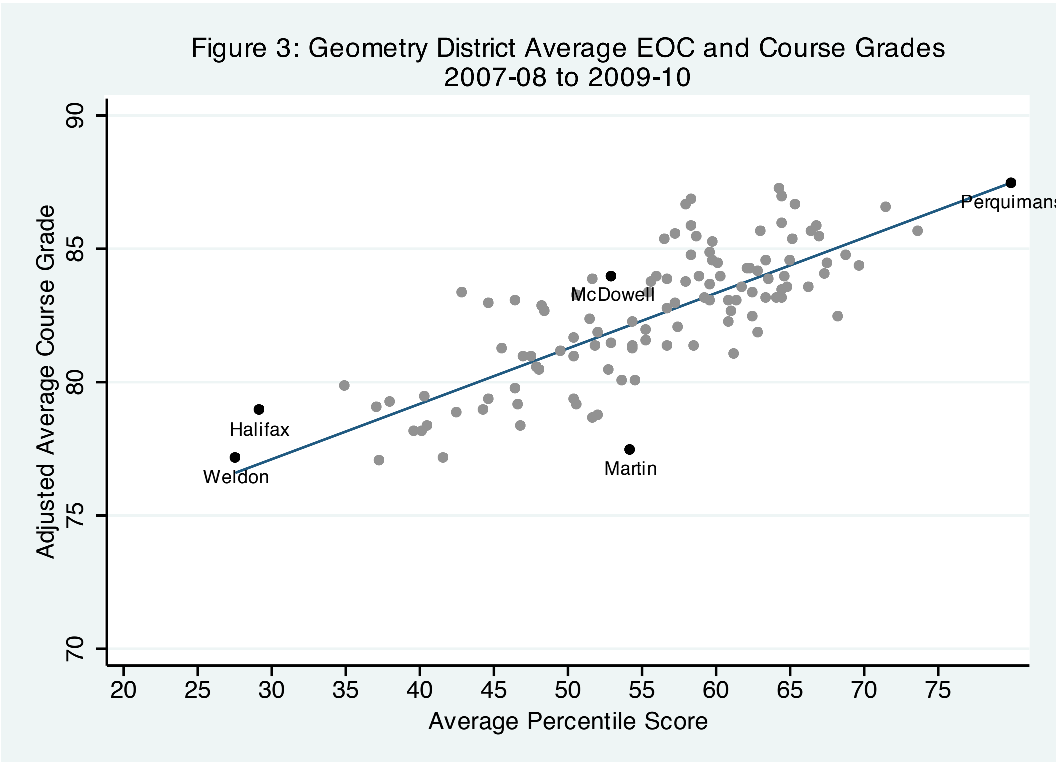
Standard 7-Point Scale			Plus/Minus 7-Point Scale		
Letter Grade	Range	Imputed Value	Letter Grade	Range	Imputed Value
			A+	99-100	99.00
A	93-100	95.45	A	96-98	97.00
			A-	93-95	94.00
			B+	91-92	91.50
B	85-92	88.29	B	88-90	89.00
			B-	85-87	86.00
			C+	83-84	83.50
C	77-84	80.59	C	80-82	81.00
			C-	77-79	78.00
			D+	75-76	75.50
D	70-76	72.65	D	73-74	73.50
			D-	70-72	71.00
F	60-69	62.49	F	60-69	62.49

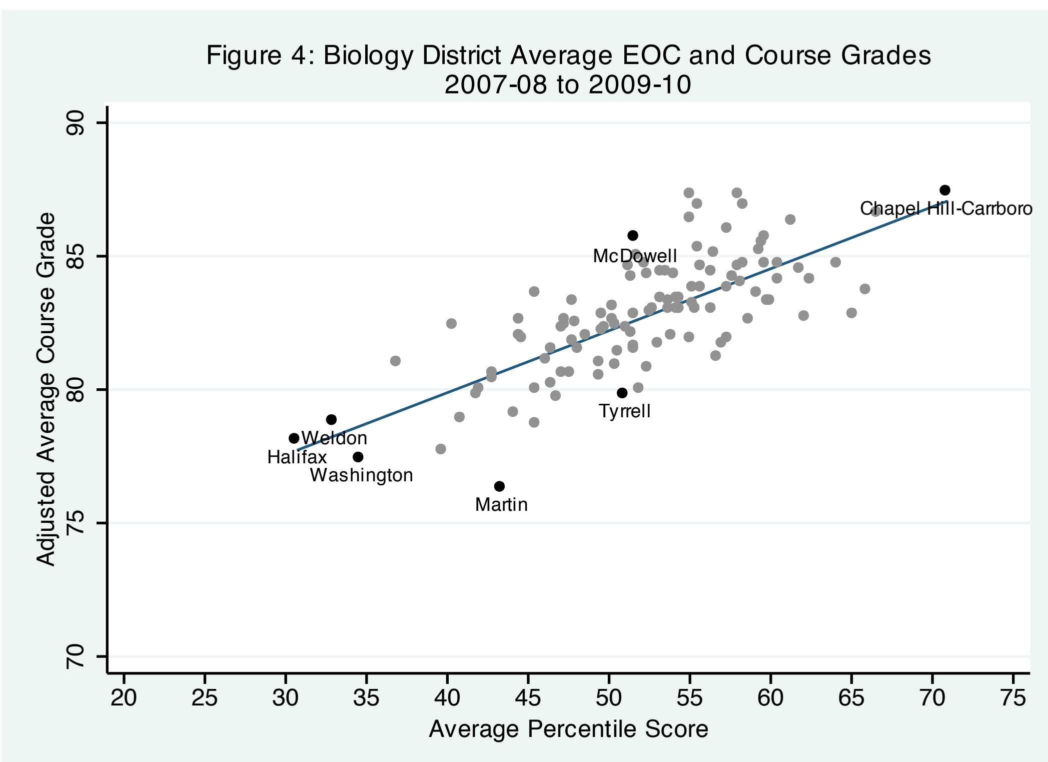
Another obstacle with the data was accounting for students who take a course in two parts. For example, some students take Algebra IA first semester and IB second semester. In these cases, I used only the grade from Algebra IB because students take the EOC test after this course, which means that the EOC test score will be included only in the grade for Algebra IB.

Some students took a course or an EOC test multiple times, sometimes within the same year. To address this issue, I only included students who had an EOC score and a course grade in the same semester. I excluded students with either a missing test score or a missing course grade. If a student had multiple test scores in the same semester, I used the highest test score. Students who receive a Level I or II score are provided with the opportunity to retake the test several days after initial testing. Teachers are most likely to input the highest score as part of the student's overall course grade, so it is the most accurate representation of the relationship between course grades and test scores.

In the teacher-level characteristics, less than 1 percent of the class size variable had a value larger than 50 students. To prevent biased findings due to these outliers, I changed all class size values above 50 to equal 50.

APPENDIX 2: Geometry, Biology, and Civics District-Level Scatterplots





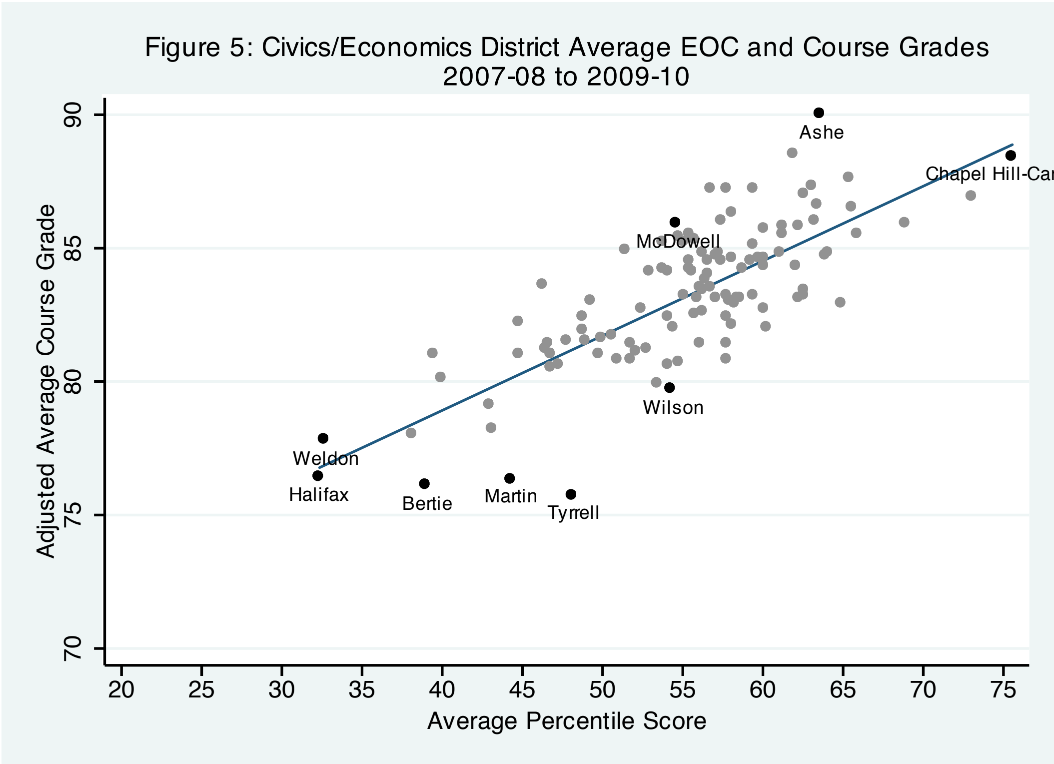


Figure 5: Civics/Economics District Average EOC and Course Grades
2007-08 to 2009-10
Adjusted Average Course Grade
Average Percentile Score
70
75
80
85
90
20
25
30
35
40
45
50
55
60
65
70
75
Ashe
Chapel Hill-Car
McDowell
Wilson
Weldon
Halifax
Bertie
Martin
Tyrrell

APPENDIX 3: Independent Variable Summary Statistics by School and Class

Table A3 provides summary statistics for school and district independent variables in all five courses.

Table A3: School and District Independent Variable Summary Statistics

Independent Variable	Mean	SD	Min	Max
School Characteristics				
Percent Poverty	39.35	17.83	0	100
Percent Black	30.75	21.91	0	98
Percent Hispanic	7.70	5.81	0	46
Percent Indian	1.28	6.13	0	84
Missed AYP*	0.88	0.32	0	1
School size (students)	1280.54	594.92	18	2948
District Characteristics				
Charlotte-Mecklenburg	0.08	0.27	0	1
Winston Salem/Forsyth	0.04	0.19	0	1
Guilford	0.05	0.22	0	1
Cumberland	0.04	0.19	0	1
Rural Coast	0.06	0.24	0	1
Rural Mountains	0.14	0.35	0	1
Rural Piedmont	0.22	0.41	0	1
Urban Coast	0.10	0.29	0	1
Urban Mountains	0.07	0.26	0	1
Urban Piedmont	0.10	0.30	0	1
Numeric 1-10 Grading Scale	0.06	0.24	0	1
7-point Letter Grade (A-F)	0.42	0.49	0	1
7-point Letter Grade (+/- A-F)	0.06	0.23	0	1
Years				
2008-09	0.35	0.48	0	1
2009-10	0.34	0.47	0	1

N=488 schools in 115 districts

Tables A4-A8 provide summary statistics for the student and teacher covariates in each of the 5 subjects. In addition, each table provides a comparison of student characteristics between the full data and the subset of data that includes all students with a matched teacher. For most covariates, the average values between groups were statistically different due to the large number of observations, even if the values were almost equivalent.

Table A4: 2007-08 to 2009-10 Algebra I Independent Variable Summary Statistics

	Full Population N=239,578				Restricted Teacher Sample N=186,711			
Independent Variable	Mean	SD	Min	Max	Mean	SD	Min	Max
Student Characteristics								
10th Grade	0.197	0.398	0	1	0.200	0.400	0	1
11th Grade	0.051	0.219	0	1	0.052	0.221	0	1
12th grade	0.012	0.109	0	1	0.012	0.111	0	1
Female	0.489	0.500	0	1	0.487	0.500	0	1
Black	0.337	0.473	0	1	0.333	0.471	0	1
Hispanic	0.095	0.293	0	1	0.097	0.296	0	1
Other	0.043	0.203	0	1	0.041	0.198	0	1
Poverty	0.611	0.487	0	1	0.610	0.488	0	1
Exceptionality	0.123	0.328	0	1	0.123	0.328	0	1
Limited English Proficiency	0.063	0.242	0	1	0.065	0.246	0	1
Teacher/Class Characteristics								
Female					0.702	0.457	0	1
Black					0.147	0.354	0	1
Hispanic					0.007	0.083	0	1
Other Race					0.012	0.107	0	1
Temporary License					0.062	0.242	0	1
Type 1 License					0.133	0.339	0	1
Masters					0.282	0.450	0	1
Class size					23.143	6.061	1	50

Table A5: 2007-08 to 2009-10 English I Independent Variable Summary Statistics

	Full Population N=305,875				Restricted Teacher Sample N=219,983			
Independent Variable	Mean	SD	Min	Max	Mean	SD	Min	Max
Student Characteristics								
10th Grade	0.024	0.153	0	1	0.026	0.158	0	1
11th Grade	0.003	0.057	0	1	0.003	0.056	0	1
12th grade	0.001	0.028	0	1	0.001	0.026	0	1
Female	0.491	0.500	0	1	0.488	0.500	0	1
Black	0.287	0.452	0	1	0.285	0.451	0	1
Hispanic	0.087	0.282	0	1	0.089	0.285	0	1
Other	0.042	0.202	0	1	0.041	0.197	0	1
Poverty	0.538	0.499	0	1	0.545	0.498	0	1
Exceptionality	0.106	0.308	0	1	0.106	0.308	0	1
Limited English Proficiency	0.055	0.227	0	1	0.056	0.231	0	1
Teacher/Class Characteristics								
Female					0.824	0.380	0	1
Black					0.135	0.341	0	1
Hispanic					0.005	0.070	0	1
Other Race					0.007	0.081	0	1
Temporary License					0.057	0.232	0	1
Type 1 License					0.140	0.347	0	1
Masters					0.324	0.468	0	1
Class size					23.132	5.960	1	50

Table A6: 2007-08 to 2009-10 Geometry Independent Variable Summary Statistics

	Full Population N=213,780				Restricted Teacher Sample N=172,294			
Independent Variable	Mean	SD	Min	Max	Mean	SD	Min	Max
Student Characteristics								
9th Grade	0.315	0.465	0	1	0.303	0.460	0	1
11th Grade	0.142	0.349	0	1	0.144	0.351	0	1
12th grade	0.021	0.142	0	1	0.021	0.144	0	1
Female	0.528	0.499	0	1	0.527	0.499	0	1
Black	0.271	0.444	0	1	0.267	0.442	0	1
Hispanic	0.069	0.254	0	1	0.070	0.255	0	1
Other	0.038	0.190	0	1	0.037	0.188	0	1
Poverty	0.443	0.497	0	1	0.446	0.497	0	1
Exceptionality	0.049	0.217	0	1	0.049	0.216	0	1
Limited English Proficiency	0.036	0.186	0	1	0.036	0.187	0	1
Teacher/Class Characteristics								
Female					0.712	0.453	0	1
Black					0.094	0.292	0	1
Hispanic					0.006	0.075	0	1
Other Race					0.010	0.099	0	1
Temporary License					0.046	0.209	0	1
Type 1 License					0.115	0.318	0	1
Masters					0.314	0.464	0	1
Class size					24.697	5.683	1	50

Table A7: 2007-08 to 2009-10 Biology Independent Variable Summary Statistics

	Full Population N=278,558				Restricted Teacher Sample N=224,830			
Independent Variable	Mean	SD	Min	Max	Mean	SD	Min	Max
Student Characteristics								
9th Grade	0.168	0.374	0	1	0.146	0.353	0	1
11th Grade	0.163	0.369	0	1	0.166	0.372	0	1
12th grade	0.037	0.189	0	1	0.038	0.192	0	1
Female	0.505	0.500	0	1	0.504	0.500	0	1
Black	0.290	0.454	0	1	0.290	0.454	0	1
Hispanic	0.075	0.263	0	1	0.076	0.265	0	1
Other	0.038	0.191	0	1	0.037	0.189	0	1
Poverty	0.494	0.500	0	1	0.502	0.500	0	1
Exceptionality	0.086	0.281	0	1	0.087	0.281	0	1
Limited English Proficiency	0.043	0.203	0	1	0.044	0.205	0	1
Teacher/Class Characteristics								
Female					0.732	0.443	0	1
Black					0.104	0.305	0	1
Hispanic					0.005	0.074	0	1
Other Race					0.014	0.119	0	1
Temporary License					0.092	0.289	0	1
Type 1 License					0.128	0.334	0	1
Masters					0.391	0.488	0	1
Class size					23.336	5.915	1	50

Table A8: 2007-08 to 2009-10 Civics Independent Variable Summary Statistics

	Full Population N= 282,702				Restricted Teacher Sample N=210,839			
Independent Variable	Mean	SD	Min	Max	Mean	SD	Min	Max
Student Characteristics								
9th Grade	0.087	0.281	0	1	0.069	0.253	0	1
11th Grade	0.091	0.288	0	1	0.095	0.293	0	1
12th grade	0.044	0.204	0	1	0.045	0.207	0	1
Female	0.505	0.500	0	1	0.501	0.500	0	1
Black	0.290	0.454	0	1	0.291	0.454	0	1
Hispanic	0.076	0.266	0	1	0.079	0.270	0	1
Other	0.038	0.192	0	1	0.038	0.191	0	1
Poverty	0.497	0.500	0	1	0.511	0.500	0	1
Exceptionality	0.089	0.284	0	1	0.091	0.288	0	1
Limited English Proficiency	0.044	0.206	0	1	0.046	0.210	0	1
Teacher/Class Characteristics								
Female					0.413	0.492	0	1
Black					0.079	0.270	0	1
Hispanic					0.003	0.057	0	1
Other Race					0.009	0.095	0	1
Temporary License					0.030	0.171	0	1
Type 1 License					0.135	0.341	0	1
Masters					0.286	0.452	0	1
Class size					23.562	6.049	1	50

APPENDIX 4: Full Regression Results for Algebra I, English I, Biology, Geometry, and Civics and Economics Regression Results

Table A9: Algebra I Regression Results

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
	β / (SE)	β / (SE)	β / (SE)	β / (SE)	β / (SE)	β / (SE)	β / (SE)
EOC Splines							
Ach Level I	0.358***	0.339***	0.348***	0.346***	0.340***	0.345***	0.340***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Ach Level II	1.242***	1.214***	1.226***	1.242***	1.268***	1.242***	1.267***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Ach Level III	0.929***	0.913***	0.916***	0.949***	0.979***	0.949***	0.979***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Ach Level IV	0.609***	0.607***	0.604***	0.638***	0.664***	0.637***	0.663***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Student Characteristics							
10th grade		-0.906***	-0.928***	-0.989***	-0.723***	-0.990***	-0.723***
		(0.10)	(0.10)	(0.11)	(0.09)	(0.11)	(0.09)
11th grade		-0.494**	-0.554***	-0.793***	-0.331**	-0.794***	-0.334**
		(0.16)	(0.16)	(0.16)	(0.12)	(0.16)	(0.12)
12th grade		1.034***	0.928***	0.703**	1.153***	0.705**	1.151***
		(0.26)	(0.27)	(0.25)	(0.21)	(0.25)	(0.21)
Female		2.292***	2.352***	2.387***	2.406***	1.997***	2.023***
		(0.04)	(0.04)	(0.05)	(0.05)	(0.07)	(0.07)
Black	0.041	0.316*	0.696***	0.774***	0.404***	0.513***	
	(0.14)	(0.14)	(0.07)	(0.07)	(0.08)	(0.06)	
Hispanic	0.512***	-0.231	0.12	0.165	-0.106	-0.019	
	(0.14)	(0.14)	(0.12)	(0.11)	(0.11)	(0.10)	
Other Race	0.023	0.211	0.158	0.117	-0.087	-0.163	
	(0.16)	(0.17)	(0.12)	(0.12)	(0.11)	(0.09)	
Black Female Interaction		-0.548***	-0.541***	-0.481***	-0.493***		
		(0.07)	(0.07)	(0.07)	(0.07)		
Hispanic Female Interaction		-0.543***	-0.527***	-0.545***	-0.485***		
		(0.11)	(0.11)	(0.12)	(0.11)		
Other Race Female Interaction		-0.512**	-0.522**	-0.516**	-0.452*		
		(0.16)	(0.16)	(0.18)	(0.18)		
Free/Reduced Price Lunch			-0.586***	-1.052***	-1.111***	-1.052***	-1.113***
			(0.08)	(0.05)	(0.04)	(0.05)	(0.04)
Exceptionality			0.687***	0.761***	0.787***	0.757***	0.783***
			(0.09)	(0.09)	(0.08)	(0.09)	(0.08)
Limited English Proficient			2.092***	2.249***	2.214***	2.257***	2.220***
			(0.14)	(0.14)	(0.12)	(0.14)	(0.12)

Table A9: Algebra I Regression Results (continued)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Teacher/Class Characteristics							
Female				-0.528***	-0.383**	-0.633***	-0.485***
				(0.14)	(0.13)	(0.14)	(0.13)
Black				0.078	0.017	-0.077	-0.028
				(0.20)	(0.20)	(0.23)	(0.24)
Hispanic				0.248	0.351	0.206	0.353
				(0.66)	(0.59)	(0.76)	(0.69)
Other Race				-0.481	-0.172	-0.343	0.053
				(0.51)	(0.50)	(0.60)	(0.52)
Temporary License				0.308	0.179	0.308	0.180
				(0.22)	(0.18)	(0.22)	(0.18)
Type 1 License				0.411*	0.170	0.414*	0.171
				(0.17)	(0.17)	(0.17)	(0.17)
Masters				0.041	-0.071	0.041	-0.071
				(0.13)	(0.13)	(0.13)	(0.13)
Class size				-0.043***	-0.045***	-0.043***	-0.045***
				(0.01)	(0.01)	(0.01)	(0.01)
Race Interactions[1]							
Black student/Black teacher						0.29	0.099
						(0.24)	(0.16)
Black student/Hispanic teacher						0.218	-0.039
						(0.57)	(0.46)
Black student/Other teacher						0.406	0.151
						(0.41)	(0.32)
Hispanic student/Black teacher						-0.14	-0.278
						(0.25)	(0.21)
Hispanic student/Hispanic teacher						-0.449	-0.036
						(0.97)	(0.88)
Hispanic student/Other teacher						-1.218	-1.115*
						(0.65)	(0.46)
Other student/Black teacher						0.229	0.488
						(0.32)	(0.25)
Other student/Hispanic teacher						0.58	0.37
						(0.88)	(0.84)
Other student/Other teacher						-1.067	-0.717
						(0.83)	(0.51)
Gender Interaction[2]							
Female student/Female teacher						0.222**	0.218**
						(0.07)	(0.07)

Table A9: Algebra I Regression Results (continued)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
School Characteristics							
Percent Poverty				0.009		0.009	
				(0.01)		(0.01)	
Percent Black				-0.004		-0.004	
				(0.01)		(0.01)	
Percent Hispanic				-0.006		-0.005	
				(0.02)		(0.02)	
Percent Indian				0.029**		0.034**	
				(0.01)		(0.01)	
Threat of missing AYP				0.174		0.182	
				(0.33)		(0.33)	
School size (students)				-0.001***		-0.001***	
				(0.00)		(0.00)	
District Characteristics							
Charlotte-Mecklenburg				0.071		0.061	
				(0.46)		(0.46)	
Cumberland				1.705*		1.691*	
				(0.77)		(0.77)	
Guilford				4.192***		4.197***	
				(0.55)		(0.55)	
Winston-Salem/Forsyth				2.633***		2.635***	
				(0.60)		(0.60)	
Rural Coast				1.351*		1.358*	
				(0.60)		(0.60)	
Rural Mountains				1.905***		1.897***	
				(0.55)		(0.55)	
Rural Piedmont				1.968***		1.959***	
				(0.51)		(0.51)	
Urban Coast				1.127*		1.131*	
				(0.56)		(0.55)	
Urban Mountains				2.042**		2.039**	
				(0.63)		(0.63)	
Urban Piedmont				1.214*		1.210*	
				(0.60)		(0.60)	
Numeric 1-10 Grade Scale	-4.253***	-4.330***	-4.330***	-4.497***		-4.496***	
	(0.29)	(0.29)	(0.29)	(0.34)		(0.34)	
7-Point Letter Grade (A-F)	-0.265	-0.269	-0.364	0.232		0.234	
	(0.26)	(0.25)	(0.26)	(0.32)		(0.32)	
7-Point Letter Grade (+/- A-F)	0.551	0.56	0.565	0.246		0.237	
	(0.37)	(0.38)	(0.39)	(0.44)		(0.44)	
Constant	17.082***	19.170***	17.902***	18.131***	19.259***	18.390***	19.536***
	(2.54)	(2.56)	(2.59)	(3.12)	(3.14)	(3.12)	(3.14)

Table A9: Algebra I Regression Results (continued)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Robust SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE?	School	School	School	School	School	School	School
Time Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School Effects	No	No	No	No	Yes	No	Yes
R^2	0.574	0.586	0.589	0.609	0.64	0.609	0.64
N	239,578	239,578	239,578	186,711	186,711	186,711	186,711
F	2112.53	1207.80	1210.12	877.12	1605.02	859.34	1361.96
p	0.000	0.000	0.000	0.000	0.000	0.000	0.000

* $p<0.05$, ** $p<0.01$, *** $p<0.001$

Dependent variable is a student's numeric course grade.

[1]Omitted Category in race interactions is white teacher/white student. Thus, coefficients represent the difference in grade in a specific group relative to this group.

[2]Omitted category in gender interactions is male teacher/male student. Thus, coefficients represent the difference in grade in a specific group relative to this group.

Table A10: English I Regression Results

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
	β / (SE)	β / (SE)	β / (SE)	β / (SE)	β / (SE)	β / (SE)	β / (SE)
EOC Splines							
Ach Level I	0.426***	0.404***	0.407***	0.416***	0.403***	0.417***	0.404***
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
Ach Level II	1.236***	1.213***	1.217***	1.242***	1.249***	1.244***	1.250***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Ach Level III	0.833***	0.795***	0.773***	0.798***	0.806***	0.798***	0.806***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Ach Level IV	0.594***	0.566***	0.540***	0.562***	0.564***	0.561***	0.562***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Student Characteristics							
10th grade		0.858***	0.854***	0.3	0.407*	0.298	0.405*
		(0.24)	(0.24)	(0.21)	(0.19)	(0.21)	(0.19)
11th grade		2.769***	2.603***	2.006***	2.005***	1.998***	2.003***
		(0.40)	(0.40)	(0.37)	(0.35)	(0.36)	(0.35)
12th grade		4.707***	4.539***	3.856***	3.542***	3.899***	3.591***
		(0.54)	(0.54)	(0.55)	(0.52)	(0.56)	(0.52)
Female		1.800***	1.849***	1.815***	1.825***	1.573***	1.583***
		(0.05)	(0.05)	(0.05)	(0.05)	(0.08)	(0.08)
Black		-1.227***	-0.649***	-0.366***	-0.320***	-0.362***	-0.308***
		(0.14)	(0.14)	(0.09)	(0.09)	(0.09)	(0.08)
Hispanic		-0.739***	-0.946***	-0.690***	-0.578***	-0.500***	-0.406***
		(0.14)	(0.16)	(0.14)	(0.13)	(0.13)	(0.12)
Other Race		-0.786**	-0.386	-0.712***	-0.659***	-0.715***	-0.630***
		(0.25)	(0.26)	(0.12)	(0.11)	(0.11)	(0.09)
Black Female Interaction		0.177*	0.183**	0.232**	0.202**		
		(0.07)	(0.07)	(0.08)	(0.07)		
Hispanic Female Interaction		0.357***	0.402***	0.421**	0.406**		
		(0.11)	(0.11)	(0.13)	(0.12)		
Other Race Female Interaction		0.007	0.005	0.096	0.084		
		(0.13)	(0.13)	(0.14)	(0.14)		
Free/Reduced Price Lunch			-1.514***	-1.904***	-1.947***	-1.902***	-1.945***
			(0.07)	(0.06)	(0.05)	(0.06)	(0.05)
Exceptionality			0.042	0.148	0.171	0.148	0.171
			(0.10)	(0.10)	(0.09)	(0.10)	(0.09)
Limited English Proficient			1.824***	2.099***	2.034***	2.091***	2.025***
			(0.15)	(0.16)	(0.14)	(0.16)	(0.14)

Table A10: English I Regression Results (continued)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Teacher/Class Characteristics							
Female				-0.129	0.058	-0.343	-0.148
				(0.18)	(0.17)	(0.19)	(0.17)
Black				0.066	0.045	-0.284	-0.248
				(0.24)	(0.22)	(0.30)	(0.27)
Hispanic				0.238	0.225	0.108	0.391
				(0.58)	(0.54)	(0.65)	(0.81)
Other Race				0.056	0.487	0.285	0.877
				(0.74)	(0.76)	(1.20)	(0.88)
Temporary License				0.443*	0.227	0.431	0.22
				(0.22)	(0.23)	(0.22)	(0.23)
Type 1 License				0.077	0.226	0.084	0.234
				(0.17)	(0.15)	(0.17)	(0.15)
Masters				-0.096	-0.02	-0.098	-0.021
				(0.14)	(0.13)	(0.14)	(0.13)
Class size				-0.029**	-0.037***	-0.028**	-0.037***
				(0.01)	(0.01)	(0.01)	(0.01)
Race Interactions[1]							
Black student/Black teacher						0.693*	0.589*
						(0.28)	(0.23)
Black student/Hispanic teacher						0.171	-0.615
						(0.96)	(0.83)
Black student/Other teacher						0.048	-0.381
						(0.85)	(0.70)
Hispanic student/Black teacher						0.165	0.218
						(0.27)	(0.23)
Hispanic student/Hispanic teacher						0.977	0.913
						(1.45)	(1.38)
Hispanic student/Other teacher						0.324	0.364
						(0.71)	(0.55)
Other student/Black teacher						0.637*	0.357
						(0.31)	(0.23)
Other student/Hispanic teacher						-1.241	-2.200
						(1.34)	(1.23)
Other student/Other teacher						-0.893	-0.954
						(1.25)	(0.93)
Gender Interaction[2]							
Female student/Female teacher						0.424***	0.411***
						(0.09)	(0.09)

Table A10: English I Regression Results (continued)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
School Characteristics							
Percent Poverty				0.021*		0.020*	
				(0.01)		(0.01)	
Percent Black				-0.004		-0.004	
				(0.01)		(0.01)	
Percent Hispanic				-0.01		-0.009	
				(0.02)		(0.02)	
Percent Indian				0.056***		0.059***	
				(0.01)		(0.01)	
Threat of missing AYP				-0.261		-0.255	
				(0.27)		(0.27)	
School size (students)				0.000		0.000	
				(0.00)		(0.00)	
District Characteristics							
Charlotte-Mecklenburg				-1.164*		-1.183*	
				(0.50)		(0.50)	
Cumberland				1.692*		1.684*	
				(0.73)		(0.73)	
Guilford				1.789**		1.762**	
				(0.66)		(0.67)	
Winston-Salem/Forsyth				1.545*		1.537*	
				(0.66)		(0.66)	
Rural Coast				0.585		0.575	
				(0.62)		(0.62)	
Rural Mountains				1.490*		1.459*	
				(0.63)		(0.63)	
Rural Piedmont				0.915		0.894	
				(0.56)		(0.55)	
Urban Coast				0.698		0.705	
				(0.56)		(0.56)	
Urban Mountains				0.519		0.49	
				(0.61)		(0.61)	
Urban Piedmont				0.781		0.768	
				(0.54)		(0.54)	
Numeric 1-10 Grade Scale	-4.639***	-4.794***	-4.835***	-4.958***		-4.951***	
	(0.26)	(0.26)	(0.26)	(0.30)		(0.30)	
7-Point Letter Grade (A-F)	-0.508*	-0.478*	-0.600**	0.079		0.078	
	(0.22)	(0.22)	(0.22)	(0.32)		(0.32)	
7-Point Letter Grade (+/- A-F)	0.566	0.606	0.63	-0.23		-0.235	
	(0.44)	(0.44)	(0.47)	(0.41)		(0.41)	
Constant	9.530**	12.438***	12.760***	11.074**	13.241***	11.080**	13.253***
	(3.08)	(3.07)	(3.07)	(3.67)	(3.59)	(3.67)	(3.59)

Table A10: English I Regression Results (continued)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Robust SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE?	School	School	School	School	School	School	School
Time Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School Effects	No	No	No	No	Yes	No	Yes
R^2	0.519	0.531	0.536	0.55	0.581	0.55	0.581
N	305,878	305,875	305,875	218,845	218,845	218,845	218,845
F	3377.14	2111.67	1897.99	1015.00	1526.88	929.72	1286.86
p	0.000	0.000	0.000	0.000	0.000	0.000	0.000

* p<0.05, ** p<0.01, *** p<0.001

Dependent variable is a student's numeric course grade.

[1]Omitted Category in race interactions is white teacher/white student. Thus, coefficients represent the difference in grade in a specific group relative to this group.

[2]Omitted category in gender interactions is male teacher/male student. Thus, coefficients represent the difference in grade in a specific group relative to this group.

Table A11: Geometry Regression Results

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
	β / (SE)	β / (SE)	β / (SE)	β / (SE)	β / (SE)	β / (SE)	β / (SE)
EOC Splines							
Ach Level I	0.623***	0.605***	0.605***	0.598***	0.594***	0.597***	0.593***
	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)
Ach Level II	0.906***	0.885***	0.882***	0.932***	0.977***	0.932***	0.978***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)
Ach Level III	1.001***	0.970***	0.966***	1.006***	1.044***	1.005***	1.043***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Ach Level IV	0.634***	0.606***	0.600***	0.637***	0.660***	0.636***	0.659***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Student Characteristics							
9th grade		0.743***	0.702***	0.624***	0.578***	0.626***	0.581***
		(0.12)	(0.12)	(0.11)	(0.09)	(0.11)	(0.09)
11th grade		-0.794***	-0.777***	-0.838***	-0.665***	-0.835***	-0.662***
		(0.12)	(0.12)	(0.11)	(0.08)	(0.11)	(0.08)
12th grade		0.216	0.215	0.066	0.144	0.065	0.145
		(0.22)	(0.22)	(0.21)	(0.15)	(0.21)	(0.15)
Female		1.668***	1.663***	1.690***	1.730***	1.464***	1.512***
		(0.04)	(0.04)	(0.04)	(0.04)	(0.07)	(0.07)
Black		-0.232	-0.02	-0.123	-0.023	-0.309***	-0.231**
		(0.15)	(0.16)	(0.08)	(0.08)	(0.09)	(0.07)
Hispanic		-0.228	-0.382**	-0.343**	-0.346***	-0.299**	-0.297***
		(0.14)	(0.14)	(0.11)	(0.10)	(0.10)	(0.09)
Other Race		-0.454**	-0.319	-0.474***	-0.397**	-0.326***	-0.290***
		(0.17)	(0.17)	(0.12)	(0.12)	(0.09)	(0.08)
Black Female Interaction		-0.311***	-0.304***	-0.255**	-0.228**		
		(0.08)	(0.08)	(0.08)	(0.07)		
Hispanic Female Interaction		0.155	0.189	0.055	0.083		
		(0.11)	(0.11)	(0.11)	(0.11)		
Other Race Female Interaction		0.331*	0.335*	0.233	0.189		
		(0.15)	(0.15)	(0.17)	(0.16)		
Free/Reduced Price Lunch			-0.515***	-0.967***	-0.990***	-0.968***	-0.989***
			(0.08)	(0.05)	(0.04)	(0.05)	(0.04)
Exceptionality			-0.501***	-0.232**	-0.182*	-0.229**	-0.179*
			(0.09)	(0.08)	(0.07)	(0.08)	(0.07)
Limited English Proficient			1.162***	1.201***	1.249***	1.204***	1.252***
			(0.13)	(0.12)	(0.10)	(0.12)	(0.10)

Table A11: Geometry Regression Results (continued)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Teacher/Class Characteristics							
Female				-0.865***	-0.883***	-0.988***	-1.012***
				(0.20)	(0.20)	(0.20)	(0.20)
Black				0.23	0.039	-0.063	-0.319
				(0.31)	(0.32)	(0.29)	(0.32)
Hispanic				-1.368	-0.556	-1.142	-0.364
				(1.23)	(0.66)	(1.19)	(0.82)
Other Race				0.221	-0.05	1.369	-0.006
				(1.13)	(0.84)	(1.31)	(0.89)
Temporary License				0.922*	0.229	0.930*	0.247
				(0.38)	(0.31)	(0.38)	(0.30)
Type 1 License				0.985***	0.830***	0.985***	0.830***
				(0.25)	(0.21)	(0.25)	(0.21)
Masters				-0.195	-0.171	-0.19	-0.16
				(0.17)	(0.18)	(0.17)	(0.18)
Class size				-0.002	-0.001	-0.002	-0.001
				(0.01)	(0.01)	(0.01)	(0.01)
Race Interactions[1]							
Black student/Black teacher						0.553	0.730**
						(0.43)	(0.26)
Black student/Hispanic teacher						-0.248	-0.435
						(1.03)	(0.76)
Black student/Other teacher						-1.318	0.368
						(0.91)	(1.01)
Hispanic student/Black teacher						0.249	0.252
						(0.33)	(0.29)
Hispanic student/Hispanic teacher						-1.221	-0.526
						(1.13)	(0.70)
Hispanic student/Other teacher						-2.024**	-0.761
						(0.69)	(0.47)
Other student/Black teacher						0.223	0.356
						(0.36)	(0.30)
Other student/Hispanic teacher						-0.36	-0.718
						(1.22)	(1.21)
Other student/Other teacher						-2.261	-1.099
						(1.32)	(0.97)
Gender Interaction[2]							
Female student/Female teacher						0.239**	0.238**
						(0.08)	(0.07)

Table A11: Geometry Regression Results (continued)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
School Characteristics							
Percent Poverty				-0.004		-0.004	
				(0.01)		(0.01)	
Percent Black				0.020**		0.019*	
				(0.01)		(0.01)	
Percent Hispanic				0.024		0.025	
				(0.02)		(0.02)	
Percent Indian				0.038*		0.040**	
				(0.02)		(0.01)	
Threat of missing AYP				0.201		0.208	
				(0.30)		(0.30)	
School size (students)				-0.001***		-0.001***	
				(0.00)		(0.00)	
District Characteristics							
Charlotte-Mecklenburg				0.501		0.495	
				(0.49)		(0.49)	
Cumberland				1.591*		1.586*	
				(0.79)		(0.79)	
Guilford				2.725***		2.734***	
				(0.68)		(0.68)	
Winston-Salem/Forsyth				3.166***		3.165***	
				(0.68)		(0.69)	
Rural Coast				0.987		0.995	
				(0.75)		(0.75)	
Rural Mountains				2.323***		2.311***	
				(0.63)		(0.63)	
Rural Piedmont				2.056***		2.056***	
				(0.60)		(0.60)	
Urban Coast				1.199		1.215	
				(0.63)		(0.63)	
Urban Mountains				2.272**		2.265**	
				(0.77)		(0.77)	
Urban Piedmont				1.339*		1.326*	
				(0.66)		(0.66)	
Numeric 1-10 Grade Scale	-5.111***	-5.179***	-5.177***	-5.515***		-5.509***	
	(0.36)	(0.35)	(0.35)	(0.38)		(0.38)	
7-Point Letter Grade (A-F)	-0.752**	-0.805**	-0.842**	-0.177		-0.172	
	(0.28)	(0.28)	(0.28)	(0.38)		(0.38)	
7-Point Letter Grade (+/- A-F)	-0.548	-0.508	-0.496	-0.668		-0.658	
	(0.46)	(0.45)	(0.46)	(0.52)		(0.52)	
Constant	-16.781***	-14.917***	-14.700***	-14.967**	-14.303**	-14.694**	-14.044**
	(4.28)	(4.33)	(4.33)	(5.23)	(5.20)	(5.22)	(5.19)

Table A11: Geometry Regression Results (continued)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Robust SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE?	School	School	School	School	School	School	School
Time Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School Effects	No	No	No	No	Yes	No	Yes
R^2	0.609	0.618	0.619	0.645	0.688	0.646	0.688
N	213,780	213,780	213,780	172,291	172,291	172,291	172,291
F	1570.47	951.35	848.42	720.33	1350.04	673.02	1178.98
p	0.000	0.000	0.000	0.000	0.000	0.000	0.000

* p<0.05, ** p<0.01, *** p<0.001
Dependent variable is a student's numeric course grade.
[1]Omitted Category in race interactions is white teacher/white student. Thus, coefficients represent the difference in grade in a specific group relative to this group.

[2]Omitted category in gender interactions is male teacher/male student. Thus, coefficients represent the difference in grade in a specific group relative to this group.

Table A12: Biology Regression Results

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
	β / (SE)	β / (SE)	β / (SE)	β / (SE)	β / (SE)	β / (SE)	β / (SE)
EOC Splines							
Ach Level I	0.584***	0.551***	0.545***	0.550***	0.534***	0.548***	0.533***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Ach Level II	1.053***	1.019***	1.001***	1.025***	1.025***	1.025***	1.025***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Ach Level III	0.843***	0.828***	0.811***	0.838***	0.851***	0.838***	0.851***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Ach Level IV	0.551***	0.552***	0.541***	0.567***	0.588***	0.566***	0.588***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Student Characteristics							
9th grade		-0.031	-0.111	0.049	-0.178	0.053	-0.176
		(0.13)	(0.13)	(0.12)	(0.12)	(0.12)	(0.12)
11th grade		-0.299*	-0.231	-0.414**	-0.596***	-0.412**	-0.595***
		(0.14)	(0.14)	(0.13)	(0.08)	(0.13)	(0.08)
12th grade		0.183	0.26	0.022	0.050	0.027	0.051
		(0.19)	(0.19)	(0.17)	(0.13)	(0.17)	(0.13)
Female		2.520***	2.476***	2.528***	2.537***	2.244***	2.265***
		(0.05)	(0.05)	(0.05)	(0.05)	(0.07)	(0.07)
Black		-0.850***	-0.473***	-0.143	-0.064	-0.391***	-0.268***
		(0.14)	(0.14)	(0.08)	(0.08)	(0.08)	(0.08)
Hispanic		-0.525***	-0.560***	-0.273*	-0.198	-0.266*	-0.223*
		(0.14)	(0.15)	(0.12)	(0.11)	(0.11)	(0.10)
Other Race		-0.611**	-0.34	-0.379***	-0.364***	-0.448***	-0.394***
		(0.23)	(0.23)	(0.11)	(0.11)	(0.09)	(0.08)
Black Female Interaction		-0.300***	-0.290***	-0.281***	-0.286***		
		(0.07)	(0.07)	(0.08)	(0.07)		
Hispanic Female Interaction		-0.058	-0.019	-0.073	-0.128		
		(0.10)	(0.10)	(0.12)	(0.11)		
Other Race Female Interaction		-0.219	-0.227	-0.266	-0.233		
		(0.14)	(0.13)	(0.15)	(0.15)		
Free/Reduced Price Lunch			-1.014***	-1.435***	-1.446***	-1.434***	-1.445***
			(0.08)	(0.05)	(0.04)	(0.05)	(0.04)
Exceptionality			-0.932***	-0.738***	-0.693***	-0.740***	-0.694***
			(0.09)	(0.09)	(0.08)	(0.09)	(0.08)
Limited English Proficient			1.103***	1.309***	1.319***	1.318***	1.324***
			(0.12)	(0.12)	(0.11)	(0.12)	(0.11)

Table A12: Biology Regression Results (continued)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Teacher/Class Characteristics							
Female				-0.600***	-0.500**	-0.726***	-0.616***
				(0.16)	(0.16)	(0.17)	(0.17)
Black				0.144	0.225	-0.215	0.093
				(0.27)	(0.27)	(0.28)	(0.26)
Hispanic				0.858*	0.763	0.79	1.102
				(0.40)	(0.58)	(0.43)	(0.64)
Other Race				0.927	0.907	1.085*	0.759
				(0.55)	(0.52)	(0.52)	(0.50)
Temporary License				0.524*	0.142	0.530*	0.147
				(0.21)	(0.21)	(0.21)	(0.21)
Type 1 License				0.481**	0.341*	0.482**	0.340*
				(0.17)	(0.17)	(0.17)	(0.17)
Masters				-0.072	-0.086	-0.077	-0.087
				(0.15)	(0.14)	(0.15)	(0.14)
Class size				0.014	-0.002	0.014	-0.002
				(0.01)	(0.01)	(0.01)	(0.01)
Race Interactions[1]							
Black student/Black teacher						0.766*	0.358
						(0.31)	(0.23)
Black student/Hispanic teacher						0.126	-0.621
						(0.59)	(0.71)
Black student/Other teacher						0.064	0.415
						(0.34)	(0.23)
Hispanic student/Black teacher						-0.170	-0.248
						(0.36)	(0.27)
Hispanic student/Hispanic teacher						-0.042	-1.005
						(1.00)	(0.68)
Hispanic student/Other teacher						-0.593	0.047
						(0.78)	(0.48)
Other student/Black teacher						-0.11	-0.421
						(0.38)	(0.33)
Other student/Hispanic teacher						0.773	0.713
						(0.52)	(0.53)
Other student/Other teacher						-0.907	-0.634
						(0.61)	(0.49)
Gender Interaction[2]							
Female student/Female teacher						0.257**	0.233**
						(0.08)	(0.08)

Table A12: Biology Regression Results (continued)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
School Characteristics							
Percent Poverty				0.012		0.012	
				(0.01)		(0.01)	
Percent Black				0.000		0.000	
				(0.01)		(0.01)	
Percent Hispanic				0.002		0.003	
				(0.02)		(0.02)	
Percent Indian				0.018		0.023	
				(0.02)		(0.02)	
Threat of missing AYP				0.109		0.117	
				(0.26)		(0.26)	
School size (students)				-0.001***		-0.001**	
				(0.00)		(0.00)	
District Characteristics							
Charlotte-Mecklenburg				0.309		0.311	
				(0.35)		(0.35)	
Cumberland				1.762**		1.770**	
				(0.60)		(0.60)	
Guilford				2.874***		2.870***	
				(0.55)		(0.54)	
Winston-Salem/Forsyth				1.827***		1.817***	
				(0.49)		(0.49)	
Rural Coast				1.03		1.025	
				(0.59)		(0.59)	
Rural Mountains				2.746***		2.724***	
				(0.51)		(0.51)	
Rural Piedmont				1.953***		1.946***	
				(0.46)		(0.46)	
Urban Coast				1.057*		1.071*	
				(0.51)		(0.51)	
Urban Mountains				2.387***		2.382***	
				(0.48)		(0.48)	
Urban Piedmont				1.385**		1.385**	
				(0.52)		(0.52)	
Numeric 1-10 Grade Scale	-5.073***	-5.231***	-5.248***	-5.588***		-5.579***	
	(0.37)	(0.37)	(0.37)	(0.39)		(0.39)	
7-Point Letter Grade (A-F)	-0.990***	-0.999***	-1.050***	-0.469		-0.467	
	(0.23)	(0.23)	(0.24)	(0.33)		(0.33)	
7-Point Letter Grade (+/- A-F)	-0.150	-0.139	-0.109	-0.389		-0.388	
	(0.49)	(0.48)	(0.49)	(0.51)		(0.52)	
Constant	-10.721**	-6.689	-5.257	-7.75	-4.567	-7.416	-4.232
	(4.14)	(4.10)	(4.11)	(4.80)	(4.66)	(4.82)	(4.67)

Table A12: Biology Regression Results (continued)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Robust SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE?	School	School	School	School	School	School	School
Time Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School Effects	No	No	No	No	Yes	No	Yes
R^2	0.567	0.585	0.589	0.603	0.639	0.603	0.639
N	278,558	278,558	278,558	223,341	223,341	223,341	223,341
F	3050.69	1738.12	1575.12	1079.02	1568.79	948.17	1270.58
p	0.000	0.000	0.000	0.000	0.000	0.000	0.000

* p<0.05, ** p<0.01, *** p<0.001

Dependent variable is a student's numeric course grade.

[1]Omitted Category in race interactions is white teacher/white student. Thus, coefficients represent the difference in grade in a specific group relative to this group.

[2]Omitted category in gender interactions is male teacher/male student. Thus, coefficients represent the difference in grade in a specific group relative to this group.

Table A13: Civics and Economics Regression Results

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
	β / (SE)	β / (SE)	β / (SE)	β / (SE)	β / (SE)	β / (SE)	β / (SE)
EOC Splines							
Ach Level I	0.478***	0.424***	0.417***	0.439***	0.438***	0.437***	0.436***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Ach Level II	1.384***	1.331***	1.307***	1.308***	1.312***	1.309***	1.313***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Ach Level III	0.880***	0.869***	0.847***	0.883***	0.891***	0.883***	0.890***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Ach Level IV	0.466***	0.477***	0.461***	0.491***	0.506***	0.490***	0.505***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Student Characteristics							
9th grade		-1.593***	-1.552***	-2.365***	-2.537***	-2.360***	-2.531***
		(0.29)	(0.29)	(0.21)	(0.19)	(0.21)	(0.19)
11th grade		-0.927***	-0.846***	-0.958***	-1.005***	-0.954***	-1.004***
		(0.10)	(0.10)	(0.11)	(0.07)	(0.11)	(0.07)
12th grade		0.368*	0.390*	0.291	0.189*	0.291	0.191*
		(0.18)	(0.17)	(0.22)	(0.09)	(0.22)	(0.09)
Female		2.527***	2.454***	2.547***	2.574***	2.204***	2.240***
		(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)
Black		-0.705***	-0.306*	0.039	0.086	-0.215**	-0.136
		(0.14)	(0.14)	(0.08)	(0.08)	(0.08)	(0.08)
Hispanic		-0.358**	-0.418**	-0.118	-0.118	-0.161	-0.157
		(0.13)	(0.14)	(0.12)	(0.11)	(0.11)	(0.10)
Other Race		-0.353	-0.071	-0.067	-0.113	-0.265**	-0.296***
		(0.23)	(0.24)	(0.15)	(0.15)	(0.10)	(0.09)
Black Female Interaction		-0.414***	-0.393***	-0.438***	-0.404***		
		(0.07)	(0.07)	(0.08)	(0.07)		
Hispanic Female Interaction		-0.169	-0.113	-0.123	-0.129		
		(0.11)	(0.11)	(0.12)	(0.11)		
Other Race Female Interaction		-0.332	-0.332	-0.383	-0.365		
		(0.19)	(0.19)	(0.21)	(0.21)		
Free/Reduced Price Lunch			-1.095***	-1.503***	-1.448***	-1.505***	-1.451***
			(0.07)	(0.05)	(0.04)	(0.05)	(0.04)
Exceptionality			-1.123***	-0.889***	-0.823***	-0.892***	-0.827***
			(0.09)	(0.09)	(0.08)	(0.09)	(0.08)
Limited English Proficient			1.157***	1.365***	1.434***	1.363***	1.431***
			(0.13)	(0.13)	(0.11)	(0.13)	(0.11)

Table A13: Civics and Economics Regression Results (continued)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Teacher/Class Characteristics							
Female				-0.422*	-0.527**	-0.658***	-0.764***
				(0.17)	(0.17)	(0.18)	(0.18)
Black				0.061	-0.006	-0.118	-0.006
				(0.31)	(0.35)	(0.37)	(0.42)
Hispanic				-1.098	-1.914***	0.011	-1.329***
				(0.67)	(0.20)	(0.41)	(0.21)
Other Race				-1.105	-0.843	-1.523	-1.673
				(0.76)	(0.85)	(1.09)	(1.07)
Temporary License				0.554	0.347	0.555	0.345
				(0.41)	(0.41)	(0.41)	(0.41)
Type 1 License				0.396*	0.447*	0.404*	0.448*
				(0.20)	(0.19)	(0.20)	(0.19)
Masters				-0.019	-0.151	-0.024	-0.152
				(0.18)	(0.18)	(0.18)	(0.18)
Class size				0.010	0.008	0.010	0.008
				(0.01)	(0.01)	(0.01)	(0.01)
Race Interactions[1]							
Black student/Black teacher						0.37	0.082
						(0.33)	(0.28)
Black student/Hispanic teacher						-1.503***	-1.001
						(0.42)	(0.57)
Black student/Other teacher						0.025	0.854*
						(0.78)	(0.38)
Hispanic student/Black teacher						-0.084	-0.307
						(0.35)	(0.29)
Hispanic student/Hispanic teacher						-1.854***	-1.221***
						(0.47)	(0.33)
Hispanic student/Other teacher						0.332	0.763
						(0.76)	(0.60)
Other student/Black teacher						-0.206	-0.283
						(0.35)	(0.33)
Other student/Hispanic teacher						-3.075***	-0.959*
						(0.64)	(0.48)
Other student/Other teacher						1.673	1.472*
						(1.10)	(0.65)
Gender Interaction[2]							
Female student/Female teacher						0.467***	0.467***
						(0.08)	(0.08)

Table A13: Civics and Economics Regression Results (continued)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
School Characteristics							
Percent Poverty				-0.008		-0.008	
				(0.01)		(0.01)	
Percent Black				0.017*		0.017*	
				(0.01)		(0.01)	
Percent Hispanic				-0.006		-0.006	
				(0.02)		(0.02)	
Percent Indian				0.043***		0.038**	
				(0.01)		(0.01)	
Threat of missing AYP				0.152		0.157	
				(0.32)		(0.32)	
School size (students)				-0.001*		-0.001*	
				(0.00)		(0.00)	
District Characteristics							
Charlotte-Mecklenburg				0.228		0.23	
				(0.55)		(0.55)	
Cumberland				2.845***		2.844***	
				(0.77)		(0.77)	
Guilford				3.095***		3.101***	
				(0.54)		(0.54)	
Winston-Salem/Forsyth				3.951***		3.954***	
				(0.62)		(0.62)	
Rural Coast				2.041**		2.032**	
				(0.65)		(0.65)	
Rural Mountains				3.728***		3.710***	
				(0.62)		(0.62)	
Rural Piedmont				2.644***		2.641***	
				(0.58)		(0.58)	
Urban Coast				2.134***		2.130***	
				(0.63)		(0.63)	
Urban Mountains				3.216***		3.200***	
				(0.64)		(0.64)	
Urban Piedmont				2.464***		2.452***	
				(0.53)		(0.53)	
Numeric 1-10 Grade Scale	-5.024***	-5.047***	-5.068***	-5.549***		-5.553***	
	(0.43)	(0.45)	(0.45)	(0.46)		(0.46)	
7-Point Letter Grade (A-F)	-0.774**	-0.666*	-0.736**	0.275		0.268	
	(0.26)	(0.26)	(0.26)	(0.34)		(0.34)	
7-Point Letter Grade (+/- A-F)	0.368	0.337	0.363	-0.106		-0.081	
	(0.42)	(0.42)	(0.42)	(0.48)		(0.48)	
Constant	2.649	9.826***	11.550***	5.712	7.662*	6.193	8.083**
	(2.75)	(2.76)	(2.76)	(3.26)	(3.08)	(3.27)	(3.10)

Table A13: Civics and Economics Regression Results (continued)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Robust SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE?	School	School	School	School	School	School	School
Time Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School Effects	No	No	No	No	Yes	No	Yes
R^2	0.57	0.59	0.593	0.616	0.651	0.617	0.651
N	282,702	282,702	282,702	210,400	210,400	210,400	210,400
F	2872.74	1620.32	1492.64	1477.78	2234.58	1591.27	2522.53
p	0.000	0.000	0.000	0.000	0.000	0.000	0.000

* $p<0.05$, ** $p<0.01$, *** $p<0.001$

Dependent variable is a student's numeric course grade.

[1]Omitted Category in race interactions is white teacher/white student. Thus, coefficients represent the difference in grade in a specific group relative to this group.

[2]Omitted category in gender interactions is male teacher/male student. Thus, coefficients represent the difference in grade in a specific group relative to this group.

Bibliography

Arcidiacono, P, Bayer, P, & Hizmo A (2010). Beyond Signaling and Human Capital: Education and the Revelation of Ability. *American Economic Journal: Applied Economics*, 2, no. 4, pp. 76–104.

Bishop, J. H. (2002). What is the appropriate role for student achievement standards?, Conference Series, Federal Reserve Bank of Boston, June Issue, pp. 249-289.

Camara, W. (1998). High School Grading Policies. College Board Research Note (RN-04), New York, NY: College Board.

Camara, W., Kimmel, E., Scheuneman, J., & Sawtell, E. A. (2003). Whose grades are inflated? (College Board Research Report No. 2003-4). New York, NY: College Board

Clark, C. (2009). Grading Disparities in Georgia's High School Algebra I, 9th Grade English Literature, 11th Grade English Literature, Biology, Physical Science, History, Geometry, and Economics Classes. Atlanta, Georgia: Governor's Office of Student Achievement.

Clotfelter, C.T., Ladd, H.F., Vigdor, J.L. (2008). School Segregation under Color-Blind Jurisprudence: The Case of North Carolina. *Virginia Journal of Social Policy & the Law*, 16, no. 1, pp. 46-86.

Dee, T.S. (2007). Teachers and the Gender Gaps in Student Achievement. *Journal of Human Resources*. 42, no. 3, pp. 528-554

Ehrenberg, R.G., Goldhaber, D.D., & Brewer, D.J. (1995). Do Teachers' Race, Gender, and Ethnicity Matter?: Evidence from Nels88. *National Bureau of Economic Research*. Working Paper No. w4669.

Evans, M.O (1992). An Estimate of Race and Gender Role-Model Effects in Teaching High School *The Journal of Economic Education*, 23, no. 3, pp. 209-217.

Fairlie, R. W., Hoffman, F., Oreopoulos, P. (2011). A Community College Teacher Like Me: Race and Ethnicity Interactions in the Classroom. Institute for the Study of Labor. Discussion Paper No. 5983.

Hardy, L. (1997). Grade inflation. *American School Board Journal*, 12, pp. 28–29.

Hinnerich, B.T., Höglin, & Johannesson M. (2011). Are Swedish boys discriminated against in high schools? *Economics of Education Review*, 30, no. 4, pp. 682-690

Hyde, J. & Jaffee, S. (1998). "Perspectives from social and feminist psychology," *Education Researcher*, 27, no. 5, pp. 14-16.

Koretz, D., & Berenz, M. (2001). Changes in high school grading standards in mathematics, 1982–1992. Washington, DC: The Rand Corporation.

Kuncel, N. R., Credé, M, & Thomas, L. L. (2005). The Validity of Self-Reported Grade Point Averages, Class Ranks, and Test Scores: A Meta-Analysis and Review of the Literature. *Review of Educational Research*. 75, no. 1, pp. 63-82.

Lavy, V. (2008). Do gender stereotypes reduce girls' and boys' human capital outcomes? Evidence from a natural experiment. *Journal of Public Economics*. 92, no. 10-11, pp. 2083-2105.

Lindahl, E. (2007). Comparing teachers' assessments and national test results—evidence from Sweden. The Institute for Labor Market Policy Evaluation. Working Paper 24.

Madon, S., Jussin, L., Keiper, S., Eccles, J., Smith, A., Palumbo, P. (1998). The Accuracy and Power of Gender, Social Class, and Ethnic Stereotypes: A Naturalistic Study in Person Perception. *Personality and Social Psychology Bulletin*, 24, no. 12, pp. 1304-1318.

U.S. Department of Education, Office of Educational Research and Improvement (1994). What do student grades mean? Differences across schools (Research Report OR 94-3401). Washington, DC: Author

Reyna, C. (2000). Lazy, Dumb, and Industrious: When Stereotypes Convey Attribution Information in the Classroom. *Education Psychology Review*, 12, no. 1, pp. 85-110.

Woodruff, D. J., & Ziomek, R. L. (2004a). High school grade inflation from 1991 to 2003. (ACT Research Report Series 2004-4). Iowa City, IA: ACT.

Texas Education Agency. (2009). A study of the correlation between course performance in Algebra I and Algebra I end-of-course test performance. Austin, TX.

Troxell, K. F. (2009). The efficacy of teacher grades in predicting standardized testing outcomes. Dissertation. University of Virginia. ProQuest Dissertations and Theses, Retrieved from http://search.proquest.com/docview/305010568?accountid=10598

Zinth, J. D. (2010). End-of-Course Exams. Denver, Colorado: Education Commission of the States.

Ziomek, R., & Svec, J. (1997). High school grades and achievement: Evidence of grade inflation. *NASSP Bulletin*, *81*, pp. 105–113.