



How Consistent are Course Grades? An Examination of Differential Grading

Samuel Rauschenberg
United States

Citation: Rauschenberg, S. (2014). How consistent are course grades? An examination of differential grading. *Education Policy Analysis Archives*, 22(92).
<http://dx.doi.org/10.14507/epaa.v22n92.2014>

Abstract: Differential grading occurs when students in courses with the same content and curriculum receive inconsistent grades across teachers, schools, or districts. It may be due to many factors, including differences in teacher grading standards, district grading policies, student behavior, teacher stereotypes, teacher quality, and curriculum adherence. If it occurs systematically, certain types of students may receive higher or lower grades relative to other students, despite having similar content mastery or ability. Using three years of statewide data on Algebra I and English I courses in North Carolina public high schools, I find that 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, holding test scores and student, teacher, school, and district characteristics constant. Low-income students, conversely, earn lower grades than other students, all else constant. With the exception of Algebra I low-income students, these differences are large enough to move a student one grade category on a plus/minus 7-point A-F grading scale. Black students earn higher Algebra I grades but lower English I grades than white or Asian students with the same test score, but these effect sizes are smaller than other student

characteristics. Interactions between student and teacher race and gender yielded small estimates that were not consistent between subjects.

Keywords: grading; achievement; stereotypes; educational economics; poverty; gender; race/ethnicity

¿Cuán consistente son las calificaciones de los cursos? Un examen de las calificaciones diferenciadas

Resumen: calificaciones diferenciadas se producen cuando alumnos de los cursos con el mismo contenido y currículo reciben calificaciones inconsistentes a través de docentes, escuelas o distritos. Este fenómeno puede deberse a muchos factores, incluyendo diferencias en las normas de clasificación de maestros, políticas de calificación distritales, comportamiento de los estudiantes, los estereotipos de maestros, la calidad del profesorado, así como la adherencia currículo. Si este fenómeno se produce sistemáticamente, ciertos tipos de estudiantes pueden recibir calificaciones más altas o más bajas en relación a otros estudiantes, a pesar de tener dominio del contenido o habilidades similares. El análisis de los datos de tres años en los cursos Inglés I y Álgebra I en las escuelas secundarias públicas de todo el estado de Carolina del Norte, muestra que las características de los estudiantes son predictores más fuertes de calificaciones diferenciadas comparadas con las características de profesor, escuela, o de distrito. Estudiantes mujeres, estudiantes con dominio limitado de Inglés, y que cursan 12^o grado obtienen calificaciones superiores que son estadísticamente significativas comparados con otros estudiantes, manteniendo constantes los resultados de exámenes, y las características de estudiantes, profesores, escuelas, y distritales constantes. Estudiantes de bajos ingresos, por el contrario, ganan calificaciones más bajas que otros estudiantes, manteniendo las demás características constantes. Con la excepción de los estudiantes de bajos ingresos en Álgebra I, estas diferencias son lo suficientemente grandes como para promover a un estudiante hasta un grado en una escala de calificación AF de más/menos de 7 puntos. Los estudiantes negros ganan calificaciones más altas en Álgebra I, pero más bajas en Inglés I que los estudiantes blancos o asiáticos con la misma resultados en las pruebas, pero la medida de esos efectos son más pequeñas que otras características de los estudiantes. Las interacciones entre estudiante y raza y género de los profesores produjeron pequeñas estimaciones que no fueron consistentes entre los sujetos.

Palabras clave: calificaciones; logro; estereotipos; economía de la educación; pobreza; género; raza/etnia.

Quão consistente são as notas dos cursos? Um exame das notas diferenciadas

Resumo: notas diferenciadas ocorrem quando os estudantes de cursos com o mesmo conteúdo e currículo recebem notas inconsistentes através de professores, escolas e distritos. Este fenômeno pode ser devido a muitos fatores, incluindo diferenças nos padrões de classificação do professor, as políticas distritais de avaliação, o comportamento dos alunos, os estereótipos de professores, a qualidade do ensino e pela adesão currículo. Se esse fenômeno ocorre sistematicamente, alguns alunos podem receber notas melhores ou piores em relação a outros estudantes, apesar de ter domínio do conteúdo ou habilidades semelhantes. A análise dos dados de três anos de cursos de Inglês I e Álgebra I em escolas públicas em todo o estado da Carolina do Norte, mostra que características dos alunos são preditores mais fortes das diferentes notas em comparação com as características do professor, escola ou distrito escolar. Estudantes do sexo feminino, estudantes com proficiência limitada em Inglês, e inscritos no 12^o ano obtiveram notas mais elevadas que são estatisticamente significativa em comparação com os outros alunos, mantendo constantes os resultados dos testes e características dos alunos, professores, escolas e distritais. Estudantes de

baixa renda, por outro lado, ganham notas mais baixas do que os outros alunos, mantendo outras características constantes. Com a exceção de estudantes de baixa renda em Álgebra I, estas diferenças são grandes o suficiente para promover um aluno para um grau em uma escala de classificação A-F mais/menos 7 pontos. Estudantes negros ganham notas mais altas em Álgebra I, mas em Inglês I menores do que os brancos ou asiáticos, com os mesmos resultados em testes, mas a extensão desses efeitos são menores do que outras características dos alunos. As interações entre raça e gênero e professores-alunos produziram pequenos estimativas não foram consistentes entre os sujeitos.

Palavras-chave: classificação; Eu realização; estereótipos; economia da educação; pobreza; sexo; raça / etnia.

Introduction

High school grades play an important role in college admissions. Many mid- and low-tier colleges make admissions decisions based almost exclusively on a student's GPA and SAT or ACT score. However, a student who receives an A in a course in one classroom may not have the same subject mastery as a student who receives an A in the same course in another classroom. This phenomenon, defined as differential grading in this paper, occurs when students in courses with the same content and curriculum receive inconsistent grades across teachers, schools, or districts (Godfrey, 2011). Many factors can lead to differential grading, including differences in teacher grading standards, district grading policies, student behavior, teacher stereotypes, teacher quality, and curriculum adherence.

High school teachers often have significant latitude in determining their grade distributions (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 or scholarship.

Differential grading may also occur if certain types of students exert different amounts of effort depending on teacher characteristics. For example, Dee (2007) finds that students with same-gender teachers perform better than students with opposite-gender teachers. Same-gender teachers can 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, Hoffman, and Oreopoulos (2011) find that underrepresented minority students have lower course withdrawal rates and higher pass rates when taught by a minority instructor.

Racial, gender, and other stereotypes of student performance also may influence how a teacher issues grades. Gender, ethnic, and socioeconomic stereotypes have some impact on how teachers view students (Madon, Jussin, Keiper, Eccles, Smith, & Palumbo, 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 (Hyde & Jaffree, 1998; Reyna, 2000). 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.

Ehrenberg, Goldhaber, and Brewer (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. Teachers routinely rate female and white students higher than other students, holding test scores constant. Further, teachers of a certain race have 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.

Lavy (2008) compares scores given by a student's teacher grades with scores given by an identity-blind grader on a similar test in nine high school courses in Israel. In each class, students took a teacher-graded school exam and an almost identical state exam within a few weeks that was graded blindly. Lavy finds that teacher-graded exam scores are higher than blind test scores for most tests. 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 are sensitive to teacher characteristics, such as gender, age, experience, and family size, but are 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 grade discrimination.

Hinnerich, Höglin, and Johannesson (2011) follow a similar methodology for a random sample of Swedish high school students' performance on a national Swedish language exam. A blind grader and the student's teacher scored each test, so the study uses blind and non-blind grading of the same test rather than two tests. Consequently, 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% lower than girls, they do not find evidence of discriminatory grading between genders. However, they find that blind grading yields scores that are 13% lower than non-blind grading for both genders, meaning that classroom teachers tend to grade their students more favorably than blind graders.

In an experiment with Indian students who were offered a monetary incentive for taking an exam, Hanna and Linden (2012) compare the scores of blind graders with graders who received a face sheet with each exam that included randomized student characteristics (age, gender, and caste). This strategy allows them to isolate discrimination in grading on the same exam. They find that teachers issue scores to "lower-caste" students 0.03 to 0.08 standard deviations below "high caste" students. They do not find patterns by gender or age.

These studies provide insight into the existence of differential grading on standardized tests. However, they do not examine differential grading in the assignment of course grades, which play a significant role in college admissions in the United States. 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 differential grading patterns, assuming the tests are unbiased.

Cornwell, Mustard, and Van Parys (2013) compare the performance of elementary students in the 1998-99 ECLS-K cohort on objective reading, math, and science assessments with their teacher's assessment of student mastery of each subject. They find that girls receive grades that, depending on the subject, race, and grade level, are generally 0.10 to 0.25 standard deviations higher than boys with the same test score. The differential grading is larger in math and science and is unaffected by teacher experience and education level. In addition, they find that controlling for non-cognitive skills reduces or eliminates the gender difference in all subjects.

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. 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. Lindahl 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 0.11 standard deviations in the grade difference distribution for girls and 0.23 standard deviations 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.

Godfrey (2011) compares Advanced Placement (AP) exam scores and actual course grades in five courses across five schools in the United States. 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. She regresses AP exam scores on AP course grades with school fixed effects and finds that the relationship between course grades and AP Exam scores varies widely within each subject and across 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. Statewide end-of-course tests (EOC) fit this description because all students who are enrolled in a class with a corresponding EOC test must take the exam. However, research on the relationship between EOC test scores and course grades has been limited. In 1999, the Texas Education Agency examined the correlation between Algebra I course grades and EOC test scores from a representative sample in one year. It finds that a larger proportion of minority and low-income students are promoted to the next level of math without adequate preparation relative to other students. Additionally, course grades explain 35% of the variation in EOC test scores, while ethnicity and income explain only 6% of the variation.

In a descriptive analysis, Clark (2009) matches course grades with Georgia EOC tests and finds sizable differential grading across districts in eight subjects. 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 perhaps because humanities courses have a larger emphasis on writing, which is more subjective to grade than math. Finally, the EOC failure rate exceeded the course failure rate for each course. 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.

In sum, previous research points to the existence of differential grading in high schools. However, it does not examine whether differential grading varies by school, teacher, or student characteristics. The current paper provides a more complete picture of differential grading by comparing course grades and EOC tests in multiple years across the population of North Carolina students in two courses, Algebra I and English I. Thus, the results have stronger external validity than previous studies. While the paper cannot distinguish between grade discrimination and other forms of differential grading, it does examine student, teacher, school, and district-level patterns across the state.

Overall, I find that 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 and student, teacher, and school characteristics constant. Low-income students, in contrast, earn lower grades than other students in both subjects. With the exception of Algebra I low-income students, these differences are large enough to move a student one grade category on a plus/minus 7-point A-F grading scale. Black students earn higher Algebra I grades but lower English I grades than white or Asian students with the same test score, but these effect sizes are smaller than other student characteristics. Interactions between student and teacher race and gender yielded small estimates that were not consistent between subjects.

The next section includes background information on EOC testing in North Carolina and describes the data used in this paper. The third section provides statewide descriptive statistics on the relationship between EOC test scores and course grades. The next section describes the empirical model, which is followed by a discussion of regression results. The paper ends with a brief conclusion.

North Carolina Testing Overview and Data

North Carolina End-of-Course Tests

First instituted in 1987, North Carolina EOC tests are aligned with the state's Standard Course of Study. 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. Schools are required to offer makeup testing to absent students for two weeks after the initial test administration date. The Algebra I exam has 64 operational items and 14 field test items. The English I exam has 56 operational test items and 24 field test items. All questions are multiple choice, and scoring is automated.

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 ranges from about 120 to 180, depending on the test. 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 Algebra I and English I, 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 and II are considered failing, while III and IV are considered passing. State law requires that a student's EOC test score constitute 25% of the overall course grade, giving students an immediate incentive to perform well on the test. Each district has a formula to convert the scale score to a 100-point converted score that teachers are expected to enter as 25% of each student's final grade.¹ The Department of Public Instruction (DPI) does not provide statewide recommendations or track the policies districts use to meet this requirement.

Data Description

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 Duke University's Sanford School of Public Policy. The data include student scores on the Algebra I and English I EOC tests and each student's grade in the course corresponding with the EOC test. In addition, the data include student-level information on race, free or reduced price lunch eligibility (FRL), exceptionality status, and Limited English Proficient (LEP) status. The NCERDC's unique student identifier allows student EOC test scores, course grades, and demographics to be linked between datasets and across the three years. Most districts reported course grades on a numeric 1-100 scale or letter grades on a

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

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.² Appendix A describes the methodology for converting grades to the same scale.

Using NCERDC's course membership data, I linked each student with his or her teacher in each course, enabling regressions to include teacher race, gender, certification type, and class size. However, due to linking limitations in the source data, only 72% of students in English I and 78% of students in Algebra I had a matching teacher. Thus, regression specifications that include teacher-level characteristics use only this subset of data. 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 in both courses. Finally, I supplemented the data with school- and district-level variables from NCERDC.

I restricted the data to include only North Carolina high school students. Since some districts encourage students to take Algebra I in middle school, high schools in these districts have fewer Algebra I students than English I. Many middle school students who enroll in Algebra I are high achieving relative to their peers. Since these students are not included in the data, the statewide average Algebra I EOC test score is lower than in English I. Appendix A provides a more complete description of the methodology used to clean and merge the data.

Descriptive Statistics on State and School-Level Patterns in Differential Grading

Statewide Patterns

Differential grading occurs when students in courses with the same content and curriculum receive inconsistent grades across teachers, schools, or districts. 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 test scores should be highly correlated, especially because the EOC test score makes up 25% of the course grade. Table 1 provides state-level summary statistics of this relationship over three years in Algebra I and English I. As noted above, the average Algebra I EOC test score and course grade are lower than in English I because I exclude middle school students who take Algebra I, many of whom are higher achieving than their peers. For each course, a 1 standard deviation change in course grade is about 10 points on a 100-point scale.

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 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 in Algebra I and English I. 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 grade for a C is 77, but the percentile score is 51.9 in Algebra I and 47.4 in English I.

² Each regression specification includes an indicator variable to control for the grading scale.

Table 1.

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

	Algebra I Mean (SD)	English I Mean (SD)
Average EOC Percentile	50.09 (25.31)	54.04 (27.12)
Average Course Grade	79.35 (10.00)	82.83 (9.53)
Correlation Coefficient	0.752	0.704
Number of Students	239,345	305,182
Number of Districts	115	115

Note: Calculations include 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. Thus, percentile rank over the 3-year period does not equal 50.

Source: Calculations from NCERDC transcript and EOC data.

Table 2.

2007-08 to 2009-10 EOC and Grade Summary Statistics

	Algebra I	English I
Percentile Score for "A" Students	81.4	82.9
Percentile Score for "B" Students	66.3	63.5
Percentile Score for "C" Students	51.9	47.4
Percentile Score for "D" Students	38.9	35.2
Percentile Score for "F" Students	18.1	18.0
Percentage of "A" students who failed EOC	1.0%	0.7%
Percentage of "B" students who failed EOC	7.1%	5.5%
Percentage of "C" students who failed EOC	24.6%	20.3%
Percentage of "D" students who failed EOC	51.5%	42.0%
Percentage of "F" students who failed EOC	90.5%	79.8%
Percentage of students who failed EOC	34.6%	21.1%
Percentage of students who failed course	15.6%	8.8%
Percentage point difference	19.0	12.3

Note: Calculations include 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. Thus, percentile rank over the 3-year period does not equal 50. Students who failed EOC earned a Level I or II score.

Source: Calculations from NCERDC transcript and EOC data.

Table 2 also counts the percentage of students in each letter grade category who failed the EOC. For example, in Algebra I, 1.0% of students who received an A in the course failed the EOC test. Similarly, 90.5% of students who received an F in the course also failed the EOC test. Thus, nearly 10% of students passed the EOC test but received an F in the course. In English I, roughly 10 percentage points fewer "D" and "F" students failed the EOC test than in Algebra I. The English I EOC test includes composition and textual analysis sections, both of which test reading and writing ability rather than subject specific content. Algebra I requires more content knowledge and skills, so students who do not learn coursework may be more penalized in Algebra I than in English I.

The final part of Table 2 shows the gap between the EOC and course failure rates in each course. 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 in English I. While more than a third of Algebra I high school test takers failed the EOC, only 15.6% 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.³

Empirical Model

The descriptive statistics in the previous section show that differential grading likely exists between North Carolina students. 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. To accomplish this goal, I use regression models with the following form:

$$G_{i,t} = \beta_0 + \beta_1 T_{i,t} + \beta_2 X_i + \beta_3 Z_i + \lambda_t + \varepsilon_{i,t}$$

where $G_{i,t}$ is student i 's numeric course grade from 60 to 100 in year t , $T_{i,t}$ is a vector of four linear splines of student i 's EOC scale score in year t , X_i is a vector of student and teacher characteristics, Z_i is a vector of school and district characteristics, 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.⁴ 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 is 139, Level II spline is 8, the Level III spline is 10, and Level IV spline is 3 points. A student with a score of 145 is in Level II by 6 points. As a result, his 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 between each achievement level. EOC percentile scores range from 0-100, while course grades range from 60-100. Level I and Level II scores are considered "failing" and thus have a narrow range on the course grade range (60-69). However, the possible scale scores range from 120 to 147. Level III and IV scores, on the other hand, range from 148 to 180 but with grade range from 70-100. Furthermore, while each district sets the conversion from EOC score to numeric grade for the 25% requirement, many use the Achievement Levels as letter grade breaks for the converted grades. In sum, this functional form captures the effect of test

³ The Algebra I gap between course failure and EOC failure was also larger than the gap in Biology, Civics and Economics, and Geometry, which I measured for a previous version of this research.

⁴ Boyd, Grossman, Lankford, Loeb, and Wyckoff (2008) note that student test scores are imperfect measures of ability because tests occur at one point in time. The resulting measurement error may bias coefficient estimates downward if test scores are an independent variable in the model. They argue that placing test scores as the dependent variable when there is a large number of observations can alleviate the measurement error. Other papers have examined the appropriateness of standardized test scores by regressing scores on course grades (see for example Brennan, Kim, Wenz-Gross, & Siperstein 2001). However, as noted previously, North Carolina teachers use a student's test score to determine a fixed portion of a student's course grade, regardless of whether the score accurately measures ability. Consequently, I chose to use test scores as a predictor of course grades in the model. In addition, placing course grades as the dependent variable provides the opportunity to examine the variation in a student's course grade for a change in each covariate while holding constant the test score.

scores on course grade more completely than a linear functional form by assuming that the relationship between EOC scores and course grades does not remain constant across achievement levels. In each specification, the coefficient on each spline was statistically different from the other spline coefficients.⁵

X_i represents a vector of student and teacher characteristics. The student characteristics include indicator variables for grade level, gender, race, free or reduced price lunch status (FRL), exceptionality status, and Limited English Proficiency (LEP) status. The regressions also include interactions terms between student gender and race indicator variables. The teacher characteristics include indicator variables for gender, race, license type, and degree level as well as a continuous variable for class size. One specification also includes interaction terms between student and teacher race as well as student and teacher gender to measure whether teachers of a certain race or gender differentially grade students of a certain race or gender. Table B1 in Appendix B includes a description of covariates.

Z_i is a vector of school and district characteristics. School characteristics include percent poverty, percent by race, school size, and an indicator variable for whether the school has missed Adequate Yearly Progress for federal accountability in at least one of the three years. The district characteristics include a vector of geographic indicator variables using the same method as Clotfelter, Ladd, and Vigdor (2008). In addition, it includes a vector of indicator variables signifying a district's grading scale.⁶ One specification replaces school and district characteristics with school fixed effects to hold constant unobserved differences between schools, including differences in grading policies or emphases on test-taking strategies. In this case, the remaining coefficients measure differential grading within schools. Table B1 in Appendix B includes a description of covariates.

The time effects, λ_t , hold constant all statewide differences across the three years of data. I use indicator variables for the 2008-09 and 2009-10 academic years, and the 2007-08 academic is the omitted year for reference.

Each specification includes robust standard errors that are clustered at the school level. The next section discusses results from each model. Appendix C provides descriptive statistics for the independent variables in each course.

Results

This section discusses the regression results for Algebra I and English I. The following subsections are split to match the groups of variables in the regressions. Full regression results are reported in Appendix E. Table 1 in Section 3 provides the mean and standard deviation for course grades. In both courses, a 1-point change in course grade is equivalent to roughly 0.1 standard deviations. In the Algebra I base model (Model 1), which includes the test score linear splines and the district grading scale variables, the observables explain 57.2% of the variation in course grades. Model (5) includes student and teacher characteristics along with school fixed effects for the subset of data that have teacher-student matches, which is 72% of students in English I and 78% in Algebra I. The observables explain 63.8% of the variation in course grades. In English I, Model (1) yields an R^2 of 0.513, and it increases to 0.576 in Model (5). Thus, test scores explain more of the

⁵ In Algebra I, the coefficients for the four Achievement Level splines are 0.357, 1.244, 0.929, and 0.609, respectively. In English I, the coefficients are 0.426, 1.237, 0.833, and 0.594, respectively. All are statistically significant with p-value less than 0.001.

⁶ Appendix A discusses the distribution of grading scales by district and how I converted all grades to a numeric scale from 60 to 100.

variation in course grades in Algebra I than English I. While the R^2 from Models (1) and (5) cannot be directly compared because the latter uses the subset of data, generally the other observables explain 6.3% to 6.6% of the grade variation in both courses.

Since this paper focuses on differential grading across a set of covariates, the results discussion emphasizes the change in course grades for a unit or indicator value change in each covariate rather than the model's explanatory power as a whole.

Student Characteristics

Differential grading exists between student characteristics. Table 3 provides regression results for student-level variables in Algebra I for Models (1), (2), (3), (4), and (5). Table 4 provides the same results for English I.

Table 3.

Algebra I Student Regression Results

	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)
	β / (SE)	β / (SE)	β / (SE)	β / (SE)	β / (SE)
<i>Student Characteristics</i>					
10th grade		-0.906*** (0.10)	-0.930*** (0.10)	-0.988*** (0.11)	-0.723*** (0.09)
11th grade		-0.519** (0.16)	-0.580*** (0.16)	-0.809*** (0.16)	-0.340** (0.12)
12th grade		1.009*** (0.26)	0.901*** (0.27)	0.699** (0.25)	1.166*** (0.21)
Female		2.298*** (0.04)	2.358*** (0.04)	2.392*** (0.05)	2.408*** (0.05)
Black		0.049 (0.14)	0.326* (0.14)	0.698*** (0.07)	0.774*** (0.07)
Hispanic		0.515*** (0.14)	-0.227 (0.14)	0.118 (0.12)	0.162 (0.11)
Other Race		0.032 (0.16)	0.222 (0.17)	0.163 (0.12)	0.123 (0.12)
Black Female Interaction		-0.553*** (0.07)	-0.546*** (0.07)	-0.486*** (0.07)	-0.496*** (0.07)
Hispanic Female Interaction		-0.545*** (0.11)	-0.530*** (0.11)	-0.546*** (0.12)	-0.484*** (0.11)
Other Race Female Interaction		-0.523** (0.16)	-0.533*** (0.16)	-0.528** (0.18)	-0.462* (0.18)
Free/Reduced Price Lunch			-0.591*** (0.08)	-1.049*** (0.05)	-1.110*** (0.04)
Exceptionality			0.693*** (0.09)	0.767*** (0.09)	0.793*** (0.08)

Table 3. (cont'd.)

Algebra I Student Regression Results

	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)
	β / (SE)	β / (SE)	β / (SE)	β / (SE)	β / (SE)
<i>Student Characteristics</i>					
Limited English Proficient			2.097*** (0.14)	2.253*** (0.14)	2.217*** (0.12)
R-Squared	0.572	0.584	0.587	0.607	0.638
Number of Observations	239,345	239,345	239,345	186,507	186,507
F-Statistic	2089.22	1208.52	1211.26	878.25	1605.37
P-Value	0.000	0.000	0.000	0.000	0.000

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

Note: Dependent variable is a student's numeric course grade. 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. Full results in Appendix E.

Model 1 includes test score linear splines, grading scale indicators, and year effects.

Model 2 adds student grade level, race, and gender.

Model 3 adds other student characteristics.

Model 4 adds teacher, school and district characteristics and uses only subset of data with teacher-student matches.

Model 5 replaces school and district characteristics with school fixed effects and uses only subset of data with teacher-student matches.

Table 4.

English I Student Regression Result

	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)
	β / (SE)	β / (SE)	β / (SE)	β / (SE)	β / (SE)
<i>Student Characteristics</i>					
10th grade		0.855*** (0.24)	0.851*** (0.24)	0.294 (0.22)	0.406* (0.19)
11th grade		2.731*** (0.40)	2.565*** (0.40)	1.964*** (0.37)	1.964*** (0.35)
12th grade		4.642*** (0.53)	4.473*** (0.54)	3.849*** (0.55)	3.543*** (0.52)
Female		1.801*** (0.05)	1.851*** (0.05)	1.816*** (0.05)	1.825*** (0.05)
Black		-1.229*** (0.14)	-0.652*** (0.14)	-0.368*** (0.09)	-0.321*** (0.09)
Hispanic		-0.743*** (0.14)	-0.950*** (0.16)	-0.692*** (0.14)	-0.581*** (0.13)
Other Race		-0.790** (0.25)	-0.389 (0.26)	-0.712*** (0.12)	-0.660*** (0.11)

Table 4. (cont'd.)

Algebra I Student Regression Results

	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)
	β / (SE)	β / (SE)	β / (SE)	β / (SE)	β / (SE)
<i>Student Characteristics</i>					
Black Female Interaction		0.173* (0.07)	0.178* (0.07)	0.230** (0.08)	0.200** (0.07)
Hispanic Female Interaction		0.358*** (0.11)	0.403*** (0.11)	0.420** (0.13)	0.405** (0.12)
Other Race Female Interaction		0.010 (0.13)	0.006 (0.13)	0.092 (0.14)	0.080 (0.14)
Free/Reduced Price Lunch			-1.511*** (0.07)	-1.905*** (0.06)	-1.947*** (0.05)
Exceptionality			0.035 (0.10)	0.139 (0.10)	0.164 (0.09)
Limited English Proficient			1.822*** (0.15)	2.095*** (0.16)	2.035*** (0.14)
R-Squared	0.513	0.525	0.531	0.545	0.576
Number of Observations	305,182	305,182	305,182	218,514	218,514
F-Statistic	3257.14	2049.37	1842.12	998.10	1528.67
P-Value	0.000	0.000	0.000	0.000	0.000

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

Note: Dependent variable is a student's numeric course grade. 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. Full results in Appendix E.

Model 1 includes test score linear splines, grading scale indicators, and year effects.

Model 2 adds student grade level, race, and gender.

Model 3 adds other student characteristics.

Model 4 adds teacher, school and district characteristics and uses only subset of data with teacher-student matches.

Model 5 replaces school and district characteristics with school fixed effects and uses only subset of data with teacher-student matches.

By grade level, Algebra I and English I have different grading patterns in 9th, 10th, and 11th grade. In Algebra I, 10th graders receive grades 0.723 to 0.988 points lower than 9th graders, holding all else constant. This difference is equivalent to 0.07 to 0.10 standard deviations. This finding may be due to course-taking patterns whereby most 10th grade Algebra I students are either weaker students or are retaking the course. Students 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 9th graders, many of whom are taking it for the first time. Eleventh graders receive higher and statistically different grades than 10th graders in every specification except Model (4). These values are still 0.340 to 0.809 points below 9th graders. English I students, in contrast, receive higher grades in each subsequent grade level relative to 9th graders. Students in 11th grade earn a course grade that is 1.964 to 2.731 higher than 9th graders in English I, equivalent to between 0.19 and 0.27 standard deviations.

Twelfth graders earn statistically significant higher grades than other grade levels in both English I and Algebra I. These grades are 0.07 to 0.12 standard deviations higher in Algebra I and 0.35 to 0.46 standard deviations higher in English I. Some of this increase is likely due to increased students' effort to graduate because they are closer to the end of school. Additionally, teachers may raise the grades of students who are in danger of failing so that they can graduate, and principals may pressure them to do so.

Differential grading also exists by gender and race. Female students earn grades that are 1.8 to 2.4 points (0.18 to 0.24 standard deviations) higher than male students in both courses regardless of race, holding test scores constant. This difference is large enough to move a female student up one category on a plus/minus 7-point grading scale, such as a B+ to an A-. These results are not sensitive to other covariates or school fixed effects. These findings are comparable to effect sizes in Lavy (2008) and Cornwell et al. (2013) but are larger than those in Lindahl (2007). Girls may be more conscientious students by turning in more work, having fewer discipline problems, and studying more than boys. In addition, girls may have higher non-cognitive skills than boys that lead to increased grades without a change in test scores (Cornwell et al., 2013). Further, 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 (Ehrenberg et al., 1995; Lavy, 2008).

In terms of race, opposite findings emerge for Algebra I and English I, and the effect sizes are smaller than with gender. The regressions include indicator variables for race and gender as well as interactions between race and gender. All minority male students in Algebra I have positive coefficients relative to white and Asian male students when teacher and school characteristics are included in the model, but only the coefficient for black male students is statistically significant. Black male students earn a grade about 0.774 points (0.08 standard deviations) higher than white or Asian male students, holding the test score, teacher characteristics, and school constant. In the same model, 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, but a large portion of this difference is driven by gender rather than race. 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, but the grade differences only represent 0.03 to 0.07 standard deviations when teacher and school characteristics are included in the regression. As with Algebra I, when teacher and school characteristics are added, the coefficients on the Black and Hispanic variables move in a positive direction.

Differential grading also occurs based upon other student characteristics. When school fixed effects are included in Model (5), Algebra I students who are eligible for free or reduced price lunch (FRL) receive grades that are 1.110 points (0.11 standard deviations) lower than other students. In English I, FRL students earn grades that are 1.947 points (0.20 standard deviations) lower than non-FRL students. The difference in Algebra I is only large enough to move borderline non-FRL students up one category on a plus/minus 7-point grading scale, but the English I difference is large enough to move non-FRL students up one category. One possible explanation for this finding is that non-FRL students may have more access to test preparation resources than FRL students. It is also possible that fewer FRL students have college aspirations than non-FRL students. High school

grades have a limited impact on the blue-collar labor market, so FRL students may not try to earn higher grades on the margin as long as they pass the course (Arcidiacono, Bayer, & Hizmo 2010). The result could also be due to grading discrimination against lower low-income students, as Hanna and Linden (2012) find in India with teachers issuing 0.03 to 0.08 standard deviation lower scores to “lower-caste” students relative to “high caste” students. Each of these factors could have some role in the differential grading, but this model does not allow for distinguishing between causes.

Students with an exceptionality receive higher grades in both courses, but the coefficients in English I are not statistically significant. In Model (5), exceptional students in Algebra I earn grades that are 0.793 (0.08 standard deviations) higher than other students. Exceptional students receive additional services relative to other students, which may benefit them more on course performance than on test performance. Additionally, 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 earn grades that are 2.217 (0.22 standard deviations) points higher than non-LEP peers in Algebra I and 2.035 (0.20 standard deviations) points higher than non-LEP peers in English I, all else constant. As with exceptional students, LEP students may benefit from the extra supports afforded to them in several ways. They receive more one-on-one attention, which may increase the amount of classwork they complete. They also may have help for their classwork, allowing them to earn higher grades than students with the same test score who do not receive such support. Lastly, teachers may try to compensate for the language barrier by artificially raising a student’s grade.

In sum, female students, LEP students, and 12th graders in both subjects earn grades that are at least 0.19 standard deviations higher than other students with the same test score. In addition, low-income students in English I receive grades that are 0.20 standard deviations lower than other students with the same test score. Low-income Algebra I students earn grades that are 0.11 points lower than non-FRL peers. With the exception of Algebra I FRL students, these differences are large enough to move a student one grade category on a plus/minus 7-point grading scale. These findings persist when the model controls for teacher, school, and other student characteristics. Finally, the sign and significance of the race indicators are not consistent in both subjects, and the effect sizes are smaller than other student characteristics.

Teacher Characteristics

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 0.379 points (0.04 standard deviations) lower than male teachers, holding test scores, student characteristics, and school characteristics constant. Other teacher characteristics are not statistically significant. None of the teacher characteristics are statistically significant in English I.

In both courses, the class size variable has a small but statistically significant negative coefficient. For each additional student, a student’s course grade decreases by 0.045 points in Algebra I and 0.037 points in English I, which is less than 0.01 standard deviations in both cases. Appendix E includes the full results for teacher characteristics.

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. While some differential grading patterns exist by race and gender within subjects, the findings do not hold for both subjects. In addition, many of the coefficients are not statistically significant, and the size of the differential grading is much smaller than the individual student variables (0.03 to 0.07 standard deviations).

Appendix D presents a table with calculated coefficients from the regression results and includes a discussion of the findings.

School Characteristics

Overall, school characteristics appear to play a minor role in differential grading. In English I, for a 10-percentage point increase in a school's low-income students, a student's course grade increases by 0.21 points (0.02 standard deviations). Percent poverty is not statistically significant in Algebra I.

For a 10-percentage point increase in American Indian students, student grades increase by 0.59 points (0.06 standard deviations) in English I and 0.034 (0.03 standard deviations) in Algebra I, all else constant. Other school-level race variables are not statistically significant.

In Algebra I, as school size increases by 100 students, a student's grade decreases by 0.1 points (0.01 standard deviations). While this effect is small, it is in the expected direction because larger schools may be more impersonal than smaller schools. The threat of missing Adequate Yearly Progress (AYP) is not significant in both courses.

Appendix E includes the full results for school characteristics.

District Characteristics

Differential grading also exists between the five large school districts and six regions in the state, as seen in Table 5. Wake County and Charlotte-Mecklenburg Schools have lower course grades than all other groups in both courses. In Algebra I, the grade difference in the other districts and regions ranged from 1.145 (0.11 standard deviations) in the Urban Coast to 4.220 (0.42 standard deviations) in Guilford. In other words, Guilford students earn grades that are 4.220 points higher than Wake or Charlotte-Mecklenburg students with the same test score, which is enough to move them up two grade categories on a plus/minus 7-point scale. Students in the Rural Mountains, Rural Piedmont, and Urban Mountains earn higher and statistically different grades from the other three geographic groups, but the differences are less than one point. In English I, Charlotte-Mecklenburg students earn grades that are 1.184 points lower than Wake County students, 2.872 points lower than Cumberland, and 2.958 points lower than Guilford students, all else constant.

Table 5.

District Characteristics Regression Results

	Algebra I	English I
	Model (6)	Model (6)
	β / (SE)	β / (SE)
<i>District Characteristics</i>		
Charlotte-Mecklenburg	0.063 (0.46)	-1.184* (0.50)
Cumberland	1.699* (0.78)	1.688* (0.73)
Guilford	4.220*** (0.55)	1.774** (0.67)
Winston-Salem/Forsyth	2.654*** (0.60)	1.531* (0.66)
Rural Coast	1.392* (0.60)	0.568 (0.62)

Table 5. (cont.'d)

District Characteristics Regression Results

	Algebra I	English I
	Model (6)	Model (6)
	β / (SE)	β / (SE)
<i>District Characteristics</i>		
Rural Mountains	1.930*** (0.55)	1.447* (0.63)
Rural Piedmont	1.977*** (0.51)	0.888 (0.55)
Urban Coast	1.145* (0.55)	0.704 (0.56)
Urban Mountains	2.044** (0.63)	0.505 (0.61)
Urban Piedmont	1.224* (0.60)	0.767 (0.54)

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

Note: Dependent variable is a student's numeric course grade. Model 6 includes student, school, and district characteristics and year effects. Omitted district is Wake County Schools.

Of the six geographic groupings, Rural Mountains is the only group with a positive coefficient that is statistically different from the other five geographic groupings in English I, but the grade difference is less than one point. While this analysis cannot determine the reasons for this differential grading, it is important to note that differential grading can also occur systematically at the district or region level.

Results 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 Algebra I and English I, holding test scores and student, teacher, school, and district characteristics constant. Low-income students, in contrast, earn lower grades than other students in both subjects, all else constant. With the exception of Algebra I low-income students, these differences are large enough to move a student one grade category on a plus/minus 7-point A-F grading scale. Black students earn higher Algebra I grades but lower English I grades than white or Asian students with the same test score, but these effect sizes are smaller than other student characteristics. Interactions between student and teacher race and gender yield small estimates that are not consistent between subjects.

Conclusion

Descriptive statistics show that significant variation exists between course grades and EOC scores in both Algebra I and English I. In addition, certain groups of students are systematically graded differently from other student groups, even when controlling for EOC test performance and other factors. This differential grading matters because course grades play an important role in college admissions. Students who receive artificially higher grades than other students with similar ability, content knowledge, and environment may have an advantage in college admissions.

Furthermore, these students may need additional remedial work to relearn material in college, which has costly implications for the students and the state.

This research adds to the literature on grading patterns in several ways. It is the first study to examine the relationship between course grades and scores on mandatory, curriculum-based tests for all high school students in a state over multiple years. As a result, the findings have stronger external validity than previous studies that use tests given primarily to high-achieving students or that have data for only one year. The paper also adds to the debate in the literature on whether teachers of a certain race or gender grade students of a certain race or gender differently.

One limitation of the research is that it cannot distinguish between causes of differential grading, such as varying grading standards, grade discrimination, or systematic differences in student behavior. Nonetheless, college admissions decisions often rest heavily on a student's GPA, and the existence of differential grading benefits some students at the expense of others with similar ability and content knowledge as measured by test scores.

References

- Arcidiacono, P., Bayer, P., & Hizmo, A. (2010). Beyond signaling and human capital: Education and the revelation of ability. *American Economic Journal: Applied Economics*, 2(4), 76–104.
- Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2008). Measuring effect sizes: the effect of measurement error. *National Center for Analysis of Longitudinal Data in Education Research*, (CALDER Working Paper No. 19). Retrieved from http://www.urban.org/UploadedPDF/1001257_measuring_effect_sizes.pdf
- Camara, W. (1998). *High school grading policies* (Research Notes No. RN-04). The College Board. Retrieved from <https://research.collegeboard.org/sites/default/files/publications/2012/7/researchnote-1998-4-high-school-grading-policies.pdf>
- 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: Governor's Office of Student Achievement. Retrieved from http://gosa.georgia.gov/sites/gosa.georgia.gov/files/related_files/document/EOCT_Grading_Alignment_Report.pdf
- 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(1), 46–86.
- Cornwell, C., Mustard, D. B., & Van Parys, J. (2013). Noncognitive skills and the gender disparities in test scores and teacher assessments: Evidence from primary school. *Journal of Human Resources*, 48(1), 236–264. <http://dx.doi.org/10.1353/jhr.2013.0002>
- Dee, T. D. (2007). Teachers and the gender gaps in student achievement. *Journal of Human Resources*, 42(3), 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* (NBER Working Paper No. 4669). Retrieved from <http://www.nber.org/papers/w4669>
- Evans, M. O. (1992). An estimate of race and gender role-model effects in teaching high school. *The Journal of Economic Education*, 23(3), 209–217. <http://dx.doi.org/10.1080/00220485.1992.10844754>

- 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). Retrieved from <http://ftp.iza.org/dp5983.pdf>
- Godfrey, K. E. (2011). *Investigating grade inflation and non-equivalence* (Research Report No. 2011-2). The College Board. Retrieved from <https://research.collegeboard.org/sites/default/files/publications/2012/7/researchreport-2011-2-investigating-grade-inflation-non-equivalence.pdf>
- Hanna, R. N., & Linden, L. L. (2012). Discrimination in grading. *American Economic Journal: Economic Policy*, 4(4), 146–168.
- Hardy, L. (1997). Grade inflation. *American School Board Journal*, 184(12), 28–30.
- Hinnerich, B. T., Höglin, E., & Johannesson, M. (2011). Are boys discriminated against in Swedish high schools? *Economics of Education Review*, 30(4), 682–690. <http://dx.doi.org/10.1016/j.econedurev.2011.02.007>
- Hyde, J. S., & Jaffee, S. (1998). Perspectives from social and feminist psychology. *Educational Researcher*, 27(5), 14–16. <http://dx.doi.org/10.3102/0013189X027005014>
- Lavy, V. (2008). Do gender stereotypes reduce girls' and boys' human capital outcomes? Evidence from a natural experiment. *Journal of Public Economics*, 92(10-11), 2083–2105. <http://dx.doi.org/10.1016/j.jpubeco.2008.02.009>
- Lindahl, E. (2007). Comparing teachers' assessments and national test results—evidence from Sweden. *The Institute for Labour Market Policy Evaluation* (Working Paper 2007:24). Retrieved from <http://www.ifau.se/Upload/pdf/se/2007/wp07-24.pdf>
- Madon, S., Jussim, 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(12), 1304–1318. <http://dx.doi.org/10.1177/01461672982412005>
- Reyna, C. (2000). Lazy, dumb, and industrious: When stereotypes convey attribution information in the classroom. *Educational Psychology Review*, 12(1), 85–110. <http://dx.doi.org/10.1023/A:1009037101170>
- Woodruff, D. J., & Ziomek, R. L. (2004). *High school grade inflation from 1991 to 2003* (Research Report No. 2004-4). ACT. Retrieved from http://www.act.org/research/researchers/reports/pdf/ACT_RR2004-4.pdf
- Texas Education Agency. (2009). *A study of the correlation between course performance in Algebra I and Algebra I end-of-course test performance*. Austin. Retrieved from <http://www.tea.state.tx.us/student.assessment/reports/>

Appendix A

Data Cleaning Methodology

This section discusses the methodology and assumptions 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. Outside 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. 48% of the students received a numeric grade from 0 to 100, and 52% received a letter grade. Of students with a letter grade, less than 10% received a grade on the plus/minus scale.

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	

Note: 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 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. 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.

To convert 10-point numeric grades to the 100-point numeric scale, I initially added a zero on the end of each number. However, the regression results in each course indicated that districts with this scale were issuing grades that were 4.5 to 5 points lower than other districts. Since this pattern occurred in all subjects, I added 4.5 to each numeric grade for students on this scale, which also coincides with the midpoint of 80 and 89. For example, I converted a numeric grade of 8 to 84.5. After this adjustment, the indicator variable was no longer statistically significant in either subject.

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

Another obstacle 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% 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 B

Variable Descriptions

Table B1

Independent Variables in Empirical Models

<i>Student Characteristics</i>	<i>Description</i>
10th Grade	= 1 if student <i>i</i> is in the 10th grade. 9th grade is the omitted category.
11th Grade	= 1 if student <i>i</i> is in the 11th grade. 9th grade is the omitted category.
12th Grade	= 1 if student <i>i</i> is in the 12th grade. 9th grade is the omitted category.
Female	= 1 if student <i>i</i> is female.
Black	= 1 if student <i>i</i> is black. White/Asian students are the omitted category.
Hispanic	= 1 if student <i>i</i> is Hispanic. White/Asian students are the omitted category.
Other	= 1 if student <i>i</i> is either multiracial or American Indian. White/Asian students are the omitted category.
Poverty	= 1 if student is ever eligible for free or reduced price lunch from 2006 to 2010.
Exceptionality	= 1 if student <i>i</i> was eligible for an exceptionality in any year.
Limited English Proficiency	= 1 if student <i>i</i> was classified as Limited English Proficient in any year.
<i>Teacher/Class Characteristics</i>	<i>Description</i>
Female	= 1 if teacher <i>j</i> is female.
Black	= 1 if teacher <i>j</i> is black. White/Asian teachers are the omitted category.
Hispanic	= 1 if teacher <i>j</i> is Hispanic. White/Asian teachers are the omitted category.
Other Race	= 1 if teacher <i>j</i> is either multiracial or American Indian. White/Asian teachers are the omitted category.
Temporary License	= 1 if teacher <i>j</i> has a temporary license. Type 2 license teachers (3+ year of experience) are the omitted category.
New Teacher	= 1 if teacher <i>j</i> has Type 1 license (0-2 years of experience). Type 2 license teachers (3+ years of experience) are the omitted category.
Masters	= 1 if teacher <i>j</i> has a master's degree or higher.
Class size	= number of students in student <i>i</i> 's class.
<i>School Characteristics</i>	<i>Description</i>
Percent Poverty	= percent of students in student <i>i</i> 's school who are eligible for free or reduced price lunch.
Percent Black	= percent of students in student <i>i</i> 's school who are black.
Percent Hispanic	= percent of students in student <i>i</i> 's school who are Hispanic.
Percent Indian	= percent of students in student <i>i</i> 's school who are Indian. School-level data do not have an "other" category.
Missed AYP*	= 1 if student <i>i</i> 's school missed Adequate Yearly Progress (AYP) targets at least once over the three years of data.

<i>Student Characteristics</i>	<i>Description</i>
School size (students)	= the number of students in student i's school.
<i>District Characteristics</i>	<i>Description</i>
Charlotte-Mecklenburg	= 1 if student i's school is located in Charlotte-Mecklenburg.
Winston Salem/Forsyth	= 1 if student i's school is located in Winston Salem/Forsyth.
Guilford	= 1 if student i's school is located in Guilford.
Cumberland	= 1 if student i's school is located in Cumberland.
Rural Coast	= 1 if student i's school is located in the rural coast.
Rural Mountains	= 1 if student i's school is located in the rural mountains.
Rural Piedmont	= 1 if student i's school is located in the rural piedmont.
Urban Coast	= 1 if student i's school is located in the urban coast.
Urban Mountains	= 1 if student i's school is located in the urban mountains.
Urban Piedmont	= 1 if student i's school is located in the urban piedmont.
Numeric 1-10 Grading Scale	= 1 if student i's district grades are reported on a numeric 1-10 grading scale. Districts with numeric 1-100 scales are the omitted category.
7-point Letter Grade (A-F)	= 1 if student i's district grades are reported on an A-F letter grade system. Districts with numeric 1-100 scales are the omitted category.
7-point Letter Grade (+/- A-F)	= 1 if student i's district grades are reported on an A-F letter grade system that includes +/- . Districts with numeric 1-100 scales are the omitted category.

Note: Under district characteristics, North Carolina regions are as defined in Clotfelter, Ladd, & Vigdor (2008). Wake County is the omitted category.

Appendix C

Independent Variable Summary Statistics by School and Class

Table C1 and C2 provide summary statistics for school and district independent variables in Algebra I and English.

Table C1

Algebra I School and District Independent Variable Summary Statistics

Independent Variable	Mean	SD	Min	Max
<i>School Characteristics</i>				
Percent Poverty	41.17	17.68	0	100
Percent Black	31.47	22.52	0	98
Percent Hispanic	7.79	5.98	0	46
Percent Indian	1.42	6.73	0	84
Missed AYP*	0.90	0.30	0	1
School size (students)	1239.32	565.13	21	2948
<i>District Characteristics</i>				
Charlotte-Mecklenburg	0.06	0.27	0	1
Winston Salem/Forsyth	0.03	0.19	0	1
Guilford	0.04	0.22	0	1
Cumberland	0.04	0.19	0	1
Rural Coast	0.06	0.24	0	1
Rural Mountains	0.16	0.35	0	1
Rural Piedmont	0.23	0.41	0	1
Urban Coast	0.10	0.29	0	1
Urban Mountains	0.07	0.26	0	1
<i>District Characteristics</i>				
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
<i>Years</i>				
2008-09	0.35	0.48	0	1
2009-10	0.35	0.47	0	1

Note: N=469 schools in 115 districts

Table C2
English I School and District Independent Variable Summary Statistics

Independent Variable	Mean	SD	Min	Max
<i>School Characteristics</i>				
Percent Poverty	39.33	17.77	0	100
Percent Black	30.72	21.90	0	98
Percent Hispanic	7.70	5.81	0	46
Percent Indian	1.28	6.14	0	84
Missed AYP*	0.88	0.32	0	1
School size (students)	1282.76	593.68	18	2948
<i>District Characteristics</i>				
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
<i>Years</i>				
2008-09	0.35	0.48	0	1
2009-10	0.34	0.47	0	1

Note: N=469 schools in 115 districts

Tables C3 and C4 provide summary statistics for the student and teacher covariates in Algebra I and English I. 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 C3
 2007-08 to 2009-10 Algebra I Independent Variable Summary Statistics

Independent Variable	Full Population N=239,345				Restricted Teacher Sample N=186,507			
	Mean	SD	Min	Max	Mean	SD	Min	Max
<i>Student Characteristics</i>								
10th Grade	0.197	0.398	0	1	0.200	0.400	0	1
11th Grade	0.051	0.219	0	1	0.051	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	0.122	0.328	0	1
Limited English Proficiency	0.063	0.242	0	1	0.065	0.246	0	1
<i>Teacher/Class Characteristics</i>								
Female					0.702	0.457	0	1
Black					0.147	0.354	0	1
Hispanic					0.007	0.083	0	1
Other Race					0.011	0.106	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					23.156	6.049	1	50

Table C4
 2007-08 to 2009-10 English I Independent Variable Summary Statistics

Independent Variable	Full Population N=305,875				Restricted Teacher Sample N=218,514			
	Mean	SD	Min	Max	Mean	SD	Min	Max
<i>Student Characteristics</i>								
10th Grade	0.024	0.152	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	0.057	0.231	0	1
<i>Teacher/Class Characteristics</i>								
Female					0.825	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	0.348	0	1
Masters					0.324	0.468	0	1
Class size					23.148	5.946	1	50

Appendix D

Teacher and Student Interactions

For ease of interpretation, Table D.1 includes the calculated coefficients from the regression results with white and Asian male students and white and Asian male teachers as the omitted categories. The following discussion uses calculations from Model (7), which includes school fixed effects. Full regression results are included in Appendix E.

In Algebra I, black students with black teachers earn grades that are 0.581 points (0.06 standard deviations) higher than white students with white teachers, all else constant. This finding is statistically significant. However, black students with black teachers earned grades that are slightly higher but not statistically different from black students with white teachers.

This pattern does not hold in English I. Black students with black teachers earn grades that are not statistically different from white students with white teachers. However, black students with black teachers earn grades that are 0.344 points (0.04 standard deviations) lower than black students with white teachers, which is a statistically significant difference. Furthermore, white teachers also issue statistically different lower grades to both Hispanic and “other” students. However, these differences represent a grading difference of less than 0.07 standard deviations.

The teacher and student gender interactions have more statistically significant results than the race interactions. In Algebra I, female students with female teachers earn grades that are 0.265 points (0.03 standard deviations) lower than female students with male teachers. In English I, however, female students with female teachers earn grades that are 0.264 points higher than female teachers with male teachers. Only the Algebra I result is statistically significant. Male Algebra I students with female teachers earn grades that are 0.480 points lower than male students with male teachers. In English I, male students with female teacher earn grades that are 0.147 points lower male students with male teachers. Only the Algebra I coefficient is statistically significant.

As with student and teacher race, the patterns in grading by teacher and student gender are not consistent between subjects. Thus, differential grading does not appear to occur systematically between teachers and students of a certain gender.

Table D1

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)
<i>Race Interactions</i>				
Black student/Black teacher	0.611	0.581	0.042	0.034
Black student/Hispanic teacher	0.833	0.827	-0.089	-0.533
Black student/Other teacher	0.498	0.755	-0.035	0.189
Black student/White teacher	0.400*	0.511*	-0.365*	-0.310*
Hispanic student/Black teacher	-0.350	-0.333	-0.622	-0.437
Hispanic student/Hispanic teacher	-0.343	0.298	0.582	0.897
Hispanic student/Other teacher	-1.642	-1.059	0.105	0.835
Hispanic student/White teacher	-0.106	-0.020	-0.504*	-0.410*
Other student/Black teacher	0.053	0.292	-0.362	-0.517
Other student/Hispanic teacher	0.704	0.561	-1.854	-2.441

Table D1 (cont.'d)

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)
<i>Race Interactions</i>				
Other student/Other teacher	-1.503	-0.821	-1.327	-0.710
Other student/White teacher	-0.088	-0.159	-0.718*	-0.634*
White student/Black teacher	-0.090	-0.022	-0.289	-0.250
White student/Hispanic teacher	0.216	0.354	0.101	0.390
White student/Other teacher	-0.348	0.056	0.277	0.877
White student/White teacher	Reference	Reference	Reference	Reference
<i>Gender Interactions</i>				
Female student/Female teacher	1.594*	1.761*	1.646	1.850
Female student/Male teacher	2.001*	2.026*	1.577*	1.586*
Male student/Female teacher	-0.626*	-0.480*	-0.355	-0.147
Male student/Male teacher	Reference	Reference	Reference	Reference

* Each coefficient used in calculation is significant with $p < 0.05$.

Note: 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. Standard errors are robust and clustered at the school level.

Appendix E

Full Regression Results for Algebra I and English I

Table E1
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)
<i>EOC Splines</i>							
Ach Level I	0.357*** (0.02)	0.337*** (0.02)	0.347*** (0.02)	0.344*** (0.02)	0.338*** (0.02)	0.343*** (0.02)	0.338*** (0.02)
Ach Level II	1.244*** (0.02)	1.216*** (0.02)	1.229*** (0.02)	1.245*** (0.02)	1.269*** (0.02)	1.245*** (0.02)	1.269*** (0.02)
Ach Level III	0.929*** (0.01)	0.913*** (0.01)	0.916*** (0.01)	0.949*** (0.01)	0.979*** (0.01)	0.949*** (0.01)	0.979*** (0.01)
Ach Level IV	0.609*** (0.01)	0.607*** (0.01)	0.604*** (0.01)	0.638*** (0.01)	0.664*** (0.01)	0.637*** (0.01)	0.663*** (0.01)
<i>Student Characteristics</i>							
10th grade		-0.906*** (0.10)	-0.930*** (0.10)	-0.988*** (0.11)	-0.723*** (0.09)	-0.989*** (0.11)	-0.724*** (0.09)
11th grade		-0.519** (0.16)	-0.580*** (0.16)	-0.809*** (0.16)	-0.340** (0.12)	-0.810*** (0.16)	-0.342** (0.12)
12th grade		1.009*** (0.26)	0.901*** (0.27)	0.699** (0.25)	1.166*** (0.21)	0.701** (0.25)	1.163*** (0.21)
Female		2.298*** (0.04)	2.358*** (0.04)	2.392*** (0.05)	2.408*** (0.05)	2.001*** (0.07)	2.026*** (0.07)
Black		0.049 (0.14)	0.326* (0.14)	0.698*** (0.07)	0.774*** (0.07)	0.400*** (0.08)	0.511*** (0.06)
Hispanic		0.515*** (0.14)	-0.227 (0.14)	0.118 (0.12)	0.162 (0.11)	-0.106 (0.11)	-0.020 (0.10)
Other Race		0.032 (0.16)	0.222 (0.17)	0.163 (0.12)	0.123 (0.12)	-0.088 (0.11)	-0.159 (0.09)
Black Female Interaction		-0.553*** (0.07)	-0.546*** (0.07)	-0.486*** (0.07)	-0.496*** (0.07)		
Hispanic Female Interaction		-0.545*** (0.11)	-0.530*** (0.11)	-0.546*** (0.12)	-0.484*** (0.11)		
Other Race Female Interaction		-0.523** (0.16)	-0.533*** (0.16)	-0.528** (0.18)	-0.462* (0.18)		
Free/Reduced Price Lunch			-0.591*** (0.08)	-1.049*** (0.05)	-1.110*** (0.04)	-1.049*** (0.05)	-1.112*** (0.04)
Exceptionality			0.693*** (0.09)	0.767*** (0.09)	0.793*** (0.08)	0.763*** (0.09)	0.790*** (0.08)
Limited English Proficient			2.097*** (0.14)	2.253*** (0.14)	2.217*** (0.12)	2.261*** (0.14)	2.223*** (0.12)

Table E1 (cont.'d)
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)
<i>Teacher/Class Characteristics</i>							
Female				-0.523*** (0.14)	-0.379** (0.13)	-0.626*** (0.14)	-0.480*** (0.13)
Black				0.070 (0.20)	0.017 (0.20)	-0.090 (0.23)	-0.022 (0.24)
Hispanic				0.257 (0.66)	0.353 (0.59)	0.216 (0.76)	0.354 (0.69)
Other Race				-0.471 (0.53)	-0.151 (0.51)	-0.348 (0.62)	0.056 (0.53)
Temporary License				0.315 (0.22)	0.179 (0.18)	0.315 (0.22)	0.180 (0.18)
Type 1 License				0.417* (0.17)	0.173 (0.17)	0.420* (0.17)	0.175 (0.17)
Masters				0.035 (0.13)	-0.070 (0.13)	0.035 (0.13)	-0.070 (0.13)
Class size				-0.042*** (0.01)	-0.045*** (0.01)	-0.042*** (0.01)	-0.045*** (0.01)
<i>Race Interactions¹</i>							
Black student/Black teacher						0.301 (0.24)	0.092 (0.16)
Black student/Hispanic teacher						0.217 (0.57)	-0.038 (0.46)
Black student/Other teacher						0.446 (0.42)	0.188 (0.32)
Hispanic student/Black teacher						-0.154 (0.25)	-0.291 (0.21)
Hispanic student/Hispanic teacher						-0.453 (0.97)	-0.036 (0.88)
Hispanic student/Other teacher						-1.188 (0.66)	-1.095* (0.46)
Other student/Black teacher						0.231 (0.32)	0.473 (0.25)
Other student/Hispanic teacher						0.576 (0.88)	0.366 (0.84)
Other student/Other teacher						-1.067 (0.83)	-0.718 (0.51)

Table E1 (cont.'d)
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)
<i>Gender Interaction²</i>							
Female student/Female teacher						0.219** (0.07)	0.215** (0.07)
<i>School Characteristics</i>							
Percent Poverty				0.009 (0.01)		0.008 (0.01)	
Percent Black				-0.003 (0.01)		-0.003 (0.01)	
Percent Hispanic				-0.006 (0.02)		-0.005 (0.02)	
Percent Indian				0.029** (0.01)		0.034** (0.01)	
Threat of missing AYP				0.168 (0.33)		0.176 (0.33)	
School size (students)				-0.001*** (0.00)		-0.001*** (0.00)	
<i>District Characteristics</i>							
Charlotte-Mecklenburg				0.074 (0.46)		0.063 (0.46)	
Cumberland				1.714* (0.77)		1.699* (0.78)	
Guilford				4.216*** (0.55)		4.220*** (0.55)	
Winston-Salem/Forsyth				2.652*** (0.60)		2.654*** (0.60)	
Rural Coast				1.384* (0.60)		1.392* (0.60)	
Rural Mountains				1.938*** (0.55)		1.930*** (0.55)	
Rural Piedmont				1.985*** (0.51)		1.977*** (0.51)	
Urban Coast				1.141* (0.55)		1.145* (0.55)	
Urban Mountains				2.047** (0.63)		2.044** (0.63)	
Urban Piedmont				1.228* (0.60)		1.224* (0.60)	
Numeric 1-10 Grade Scale	0.250 (0.29)	0.174 (0.29)	0.173 (0.29)	0.002 (0.34)		0.003 (0.33)	
7-Point Letter Grade (A-F)	-0.271 (0.26)	-0.275 (0.25)	-0.371 (0.26)	0.223 (0.33)		0.226 (0.33)	
7-Point Letter Grade (+/- A-F)	0.554 (0.37)	0.563 (0.38)	0.568 (0.39)	0.245 (0.44)		0.237 (0.44)	
Constant	17.248*** (2.54)	19.347*** (2.56)	18.066*** (2.59)	18.308*** (3.13)	19.848*** (3.14)	18.575*** (3.13)	20.130*** (3.15)

Table E1 (cont.'d)

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)
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	0.572	0.584	0.587	0.607	0.638	0.607	0.637
N	239,345	239,345	239,345	186,507	186,507	186,507	186,507
F	2089.22	1208.52	1211.26	878.25	1605.37	861.35	1362.05
p	0.000	0.000	0.000	0.000	0.000	0.000	0.000

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

Note: Dependent variable is a student's numeric course grade. Omitted Category in race interactions is white or Asian teacher/white or Asian student. Thus, coefficients represent the difference in grade in a specific group relative to this group. 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. Standard errors are robust and clustered at the school level. Models (4), (5), (6), and (7) use only the subset of data with teacher-student matches.

Table E2
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)
<i>EOC Splines</i>							
Ach Level I	0.426*** (0.02)	0.404*** (0.02)	0.407*** (0.02)	0.415*** (0.03)	0.403*** (0.03)	0.417*** (0.03)	0.404*** (0.03)
Ach Level II	1.237*** (0.02)	1.214*** (0.02)	1.217*** (0.02)	1.242*** (0.02)	1.250*** (0.02)	1.244*** (0.02)	1.251*** (0.02)
Ach Level III	0.833*** (0.01)	0.795*** (0.01)	0.773*** (0.01)	0.798*** (0.01)	0.806*** (0.01)	0.798*** (0.01)	0.806*** (0.01)
Ach Level IV	0.594*** (0.01)	0.566*** (0.01)	0.540*** (0.01)	0.562*** (0.01)	0.563*** (0.01)	0.561*** (0.01)	0.562*** (0.01)
<i>Student Characteristics</i>							
10th grade		0.855*** (0.24)	0.851*** (0.24)	0.294 (0.22)	0.406* (0.19)	0.292 (0.22)	0.404* (0.19)
11th grade		2.731*** (0.40)	2.565*** (0.40)	1.964*** (0.37)	1.964*** (0.35)	1.955*** (0.36)	1.962*** (0.35)
12th grade		4.642*** (0.53)	4.473*** (0.54)	3.849*** (0.55)	3.543*** (0.52)	3.891*** (0.56)	3.592*** (0.52)
Female		1.801*** (0.05)	1.851*** (0.05)	1.816*** (0.05)	1.825*** (0.05)	1.577*** (0.08)	1.586*** (0.08)
Black		-1.229*** (0.14)	-0.652*** (0.14)	-0.368*** (0.09)	-0.321*** (0.09)	-0.365*** (0.09)	-0.310*** (0.08)
Hispanic		-0.743*** (0.14)	-0.950*** (0.16)	-0.692*** (0.14)	-0.581*** (0.13)	-0.504*** (0.13)	-0.410*** (0.12)
Other Race		-0.790** (0.25)	-0.389 (0.26)	-0.712*** (0.12)	-0.660*** (0.11)	-0.718*** (0.11)	-0.634*** (0.09)
Black Female Interaction		0.173* (0.07)	0.178* (0.07)	0.230** (0.08)	0.200** (0.07)		
Hispanic Female Interaction		0.358*** (0.11)	0.403*** (0.11)	0.420** (0.13)	0.405** (0.12)		
Other Race Female Interaction		0.010 (0.13)	0.006 (0.13)	0.092 (0.14)	0.080 (0.14)		
Free/Reduced Price Lunch			-1.511*** (0.07)	-1.905*** (0.06)	-1.947*** (0.05)	-1.904*** (0.06)	-1.945*** (0.05)
Exceptionality			0.035 (0.10)	0.139 (0.10)	0.164 (0.09)	0.139 (0.10)	0.164 (0.09)
Limited English Proficient			1.822*** (0.15)	2.095*** (0.16)	2.035*** (0.14)	2.087*** (0.16)	2.026*** (0.14)
<i>Teacher/Class Characteristics</i>							
Female				-0.144 (0.18)	0.057 (0.17)	-0.355 (0.19)	-0.147 (0.17)
Black				0.063 (0.24)	0.046 (0.22)	-0.289 (0.30)	-0.250 (0.27)
Hispanic				0.233 (0.58)	0.226 (0.54)	0.101 (0.65)	0.390 (0.81)
Other Race				0.053 (0.74)	0.487 (0.76)	0.277 (1.21)	0.877 (0.88)
Temporary License				0.434 (0.22)	0.223 (0.23)	0.421 (0.22)	0.216 (0.23)
Type 1 License				0.073	0.225	0.080	0.234

Table E2 (cont'd.)
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)
<i>Teacher/Class Characteristics</i>							
Masters				-0.094 (0.14)	-0.020 (0.13)	-0.096 (0.14)	-0.020 (0.13)
Class size				-0.029** (0.01)	-0.037*** (0.01)	-0.029** (0.01)	-0.037*** (0.01)
<i>Race Interactions¹</i>							
Black student/Black teacher						0.696* (0.28)	0.594* (0.23)
Black student/Hispanic teacher						0.175 (0.96)	-0.613 (0.83)
Black student/Other teacher						0.053 (0.85)	-0.378 (0.70)
Hispanic student/Black teacher						0.171 (0.27)	0.223 (0.23)
Hispanic student/Hispanic teacher						0.985 (1.45)	0.917 (1.38)
Hispanic student/Other teacher						0.332 (0.71)	0.368 (0.55)
Other student/Black teacher						0.645* (0.31)	0.367 (0.23)
Other student/Hispanic teacher						-1.237 (1.34)	-2.197 (1.23)
Other student/Other teacher						-0.886 (1.26)	-0.953 (0.93)
<i>Gender Interaction²</i>							
Female student/Female teacher						0.418*** (0.09)	0.406*** (0.09)
<i>School Characteristics</i>							
Percent Poverty				0.021* (0.01)		0.021* (0.01)	
Percent Black				-0.004 (0.01)		-0.004 (0.01)	
Percent Hispanic				-0.011 (0.02)		-0.010 (0.02)	
Percent Indian				0.056*** (0.01)		0.059*** (0.01)	
Threat of missing AYP				-0.257 (0.28)		-0.251 (0.28)	
School size (students)				0.000 (0.00)		0.000 (0.00)	
<i>District Characteristics</i>							
Charlotte-Mecklenburg				-1.166* (0.50)		-1.184* (0.50)	

Table E2 (cont'd.)
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)
<i>District Characteristics</i>							
Cumberland				1.697* (0.73)		1.688* (0.73)	
Guilford				1.801** (0.67)		1.774** (0.67)	
Winston-Salem/Forsyth				1.539* (0.66)		1.531* (0.66)	
Rural Coast				0.578 (0.62)		0.568 (0.62)	
Rural Mountains				1.479* (0.63)		1.447* (0.63)	
Rural Piedmont				0.910 (0.56)		0.888 (0.55)	
Urban Coast				0.697 (0.56)		0.704 (0.56)	
Urban Mountains				0.533 (0.61)		0.505 (0.61)	
Urban Piedmont				0.780 (0.54)		0.767 (0.54)	
Numeric 1-10 Grade Scale	-0.144 (0.26)	-0.300 (0.26)	-0.340 (0.26)	-0.454 (0.30)		-0.447 (0.30)	
7-Point Letter Grade (A-F)	-0.506* (0.22)	-0.477* (0.22)	-0.599** (0.23)	0.092 (0.32)		0.090 (0.32)	
7-Point Letter Grade (+/- A-F)	0.568 (0.44)	0.607 (0.44)	0.631 (0.47)	-0.225 (0.41)		-0.230 (0.41)	
Constant	9.433** (3.10)	12.394*** (3.08)	12.727*** (3.09)	11.139** (3.69)	13.597*** (3.61)	11.139** (3.69)	13.607*** (3.61)
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	0.513	0.525	0.531	0.545	0.576	0.545	0.576
N	305,182	305,182	305,182	218,514	218,514	218,514	218,514
F	3257.14	2049.37	1842.12	998.10	1528.67	913.76	1288.25
p	0.000	0.000	0.000	0.000	0.000	0.000	0.000

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

Note: Dependent variable is a student's numeric course grade. Omitted Category in race interactions is white or Asian teacher/white or Asian student. Thus, coefficients represent the difference in grade in a specific group relative to this group. 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. Standard errors are robust and clustered at the school level. Models (4), (5), (6), and (7) use only the subset of data with teacher-student matches.

About the Author

Sam Rauschenberg

sam.rauschenberg@gmail.com

Sam Rauschenberg is currently the Deputy Director of the Georgia Governor's Office of Student Achievement. He holds a Master of Public Policy from the Sanford School of Public Policy at Duke University and a Bachelor of Science in Economics from Georgia College.

Acknowledgements

The author is grateful to Clara Muschkin and Kara Bonneau with the North Carolina Education Researcher Data Center (NCERDC) for providing access to the data. In addition, he would like to thank Helen Ladd for serving as the primary advisor for this project. In addition, Seth Sanders, Charles Clotfelter, Benjamin Scafidi, Christopher Clark, Jennifer Heissel, and Walker Swain provided helpful feedback and advice. The author is responsible for any remaining errors or omissions. Any views expressed in the paper do not reflect the views of Duke University or the Governor's Office of Student Achievement.

education policy analysis archives

Volume 22 Number 92

September 22nd, 2014

ISSN 1068-2341



Readers are free to copy, display, and distribute this article, as long as the work is attributed to the author(s) and **Education Policy Analysis Archives**, it is distributed for non-commercial purposes only, and no alteration or transformation is made in the work. More details of this Creative Commons license are available at <http://creativecommons.org/licenses/by-nc-sa/3.0/>. All other uses must be approved by the author(s) or **EPAA**. **EPAA** is published by the Mary Lou Fulton Institute and Graduate School of Education at Arizona State University. Articles are indexed in CIRC (Clasificación Integrada de Revistas Científicas, Spain), DIALNET (Spain), [Directory of Open Access Journals](#), EBSCO Education Research Complete, ERIC, Education Full Text (H.W. Wilson), QUALIS A2 (Brazil), SCImago Journal Rank; SCOPUS, Socolar (China).

Please contribute commentaries at <http://epaa.info/wordpress/> and send errata notes to Gustavo E. Fischman fischman@asu.edu

Join **EPAA's Facebook community** at <https://www.facebook.com/EPAAAPE> and **Twitter feed** @epaa_aape.

education policy analysis archives
editorial board

Editor **Gustavo E. Fischman** (Arizona State University)

Associate Editors: **Audrey Amrein-Beardsley** (Arizona State University) **Rick Mintrop**, (University of California, Berkeley) **Jeanne M. Powers** (Arizona State University)

Jessica Allen University of Colorado, Boulder

Gary Anderson New York University

Michael W. Apple University of Wisconsin, Madison

Angela Arzubiaga Arizona State University

David C. Berliner Arizona State University

Robert Bickel Marshall University

Henry Braun Boston College

Eric Camburn University of Wisconsin, Madison

Wendy C. Chi University of Colorado, Boulder

Casey Cobb University of Connecticut

Arnold Danzig Arizona State University

Antonia Darder University of Illinois, Urbana-Champaign

Linda Darling-Hammond Stanford University

Chad d'Entremont Strategies for Children

John Diamond Harvard University

Tara Donahue Learning Point Associates

Sherman Dorn University of South Florida

Christopher Joseph Frey Bowling Green State University

Melissa Lynn Freeman Adams State College

Amy Garrett Dikkers University of Minnesota

Gene V Glass Arizona State University

Ronald Glass University of California, Santa Cruz

Harvey Goldstein Bristol University

Jacob P. K. Gross Indiana University

Eric M. Haas WestEd

Kimberly Joy Howard* University of Southern California

Aimee Howley Ohio University

Craig Howley Ohio University

Steve Klees University of Maryland

Jackyung Lee SUNY Buffalo

Christopher Lubienski University of Illinois, Urbana-Champaign

Sarah Lubienski University of Illinois, Urbana-Champaign

Samuel R. Lucas University of California, Berkeley

Maria Martinez-Coslo University of Texas, Arlington

William Mathis University of Colorado, Boulder

Tristan McCowan Institute of Education, London

Heinrich Mintrop University of California, Berkeley

Michele S. Moses University of Colorado, Boulder

Julianne Moss University of Melbourne

Sharon Nichols University of Texas, San Antonio

Noga O'Connor University of Iowa

João Paraskveva University of Massachusetts, Dartmouth

Laurence Parker University of Illinois, Urbana-Champaign

Susan L. Robertson Bristol University

John Rogers University of California, Los Angeles

A. G. Rud Purdue University

Felicia C. Sanders The Pennsylvania State University

Janelle Scott University of California, Berkeley

Kimberly Scott Arizona State University

Dorothy Shipps Baruch College/CUNY

Maria Teresa Tatto Michigan State University

Larisa Warhol University of Connecticut

Cally Waite Social Science Research Council

John Weathers University of Colorado, Colorado Springs

Kevin Welner University of Colorado, Boulder

Ed Wiley University of Colorado, Boulder

Terrence G. Wiley Arizona State University

John Willinsky Stanford University

Kyo Yamashiro University of California, Los Angeles

archivos analíticos de políticas educativas consejo editorial

Editor: **Gustavo E. Fischman** (Arizona State University)

Editores. Asociados **Jason Beech** (Universidad de San Andrés) **Alejandro Canales** (UNAM) y **Jesús Romero Morante** (Universidad de Cantabria)

Armando Alcántara Santuario Instituto de Investigaciones sobre la Universidad y la Educación, UNAM México

Claudio Almonacid Universidad Metropolitana de Ciencias de la Educación, Chile

Pilar Arnaiz Sánchez Universidad de Murcia, España

Xavier Besalú Costa Universitat de Girona, España

Jose Joaquín Brunner Universidad Diego Portales, Chile

Damián Canales Sánchez Instituto Nacional para la Evaluación de la Educación, México

María Caridad García Universidad Católica del Norte, Chile

Raimundo Cuesta Fernández IES Fray Luis de León, España

Marco Antonio Delgado Fuentes Universidad Iberoamericana, México

Inés Dussel DIE, Mexico

Rafael Feito Alonso Universidad Complutense de Madrid, España

Pedro Flores Crespo Universidad Iberoamericana, México

Verónica García Martínez Universidad Juárez Autónoma de Tabasco, México

Francisco F. García Pérez Universidad de Sevilla, España

Edna Luna Serrano Universidad Autónoma de Baja California, México

Alma Maldonado Departamento de Investigaciones Educativas, Centro de Investigación y de Estudios Avanzados, México

Alejandro Márquez Jiménez Instituto de Investigaciones sobre la Universidad y la Educación, UNAM México

José Felipe Martínez Fernández University of California Los Angeles, USA

Fanni Muñoz Pontificia Universidad Católica de Perú

Imanol Ordorika Instituto de Investigaciones Economicas – UNAM, México

Maria Cristina Parra Sandoval Universidad de Zulia, Venezuela

Miguel A. Pereyra Universidad de Granada, España

Monica Pini Universidad Nacional de San Martín, Argentina

Paula Razquin Universidad de San Andrés

Ignacio Rivas Flores Universidad de Málaga, España

Daniel Schugurensky Arizona State University

Orlando Pulido Chaves Universidad Pedagógica Nacional, Colombia

José Gregorio Rodríguez Universidad Nacional de Colombia

Miriam Rodríguez Vargas Universidad Autónoma de Tamaulipas, México

Mario Rueda Beltrán Instituto de Investigaciones sobre la Universidad y la Educación, UNAM México

José Luis San Fabián Maroto Universidad de Oviedo, España

Yengny Marisol Silva Laya Universidad Iberoamericana, México

Aida Terrón Bañuelos Universidad de Oviedo, España

Jurjo Torres Santomé Universidad de la Coruña, España

Antoni Verger Planells University of Amsterdam, Holanda

Mario Yapu Universidad Para la Investigación Estratégica, Bolivia

arquivos analíticos de políticas educativas
conselho editorial

Editor: **Gustavo E. Fischman** (Arizona State University)
Editores Associados: **Rosa Maria Bueno Fisher** e **Luis A. Gandin**
(Universidade Federal do Rio Grande do Sul)

Dalila Andrade de Oliveira Universidade Federal de Minas Gerais, Brasil
Paulo Carrano Universidade Federal Fluminense, Brasil
Alicia Maria Catalano de Bonamino Pontifícia Universidade Católica-Rio, Brasil
Fabiana de Amorim Marcello Universidade Luterana do Brasil, Canoas, Brasil
Alexandre Fernandez Vaz Universidade Federal de Santa Catarina, Brasil
Gaudêncio Frigotto Universidade do Estado do Rio de Janeiro, Brasil
Alfredo M Gomes Universidade Federal de Pernambuco, Brasil
Petronilha Beatriz Gonçalves e Silva Universidade Federal de São Carlos, Brasil
Nadja Herman Pontifícia Universidade Católica –Rio Grande do Sul, Brasil
José Machado Pais Instituto de Ciências Sociais da Universidade de Lisboa, Portugal
Wenceslao Machado de Oliveira Jr. Universidade Estadual de Campinas, Brasil

Jefferson Mainardes Universidade Estadual de Ponta Grossa, Brasil
Luciano Mendes de Faria Filho Universidade Federal de Minas Gerais, Brasil
Lia Raquel Moreira Oliveira Universidade do Minho, Portugal
Belmira Oliveira Bueno Universidade de São Paulo, Brasil
Antônio Teodoro Universidade Lusófona, Portugal
Pia L. Wong California State University Sacramento, U.S.A
Sandra Regina Sales Universidade Federal Rural do Rio de Janeiro, Brasil
Elba Siqueira Sá Barreto Fundação Carlos Chagas, Brasil
Manuela Terrasêca Universidade do Porto, Portugal
Robert Verhine Universidade Federal da Bahia, Brasil
Antônio A. S. Zuin Universidade Federal de São Carlos, Brasil