THE IMPACT OF DIFFERENTIAL GRADING STANDARDS ON STUDENT ACHIEVEMENT

by Kenneth Thomas Whelan

THE IMPACT OF DIFFERENTIAL GRADING STANDARDS ON STUDENT ACHIEVEMENT

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

by

Kenneth Thomas Whelan

August 2010

ABSTRACT

In this paper I propose and evaluate a grading standard design that allows for multiple grading standards across the grade distribution. I show that making it harder for a student to receive a higher grade when at a higher point in the grade distribution than to receive a higher grade when at some point lower leads to larger test score gains overall. Effects differ by race, gender, and at different points in the achievement distribution. Much of this analysis follows the method used by Betts and Grogger (2003) and I present my replications of their results for comparison purposes.

BIOGRAPHICAL SKETCH

Ken Whelan earned Bachelors of Science degrees in Accountancy and Finance in 2004 from Arizona State University. He is a Certified Public Accountant in the state of Massachusetts.

ACKNOWLEDGEMENTS

I would especially like to thank Ronald Ehrenberg and Jordan Matsudaira for their continued help with this project. I would also like to thank C. Kirabo Jackson, Ben Ost, Doug Webber, Joyce Main, Mirinda Martin, Joshua Price, and Jenny Lewis for helpful comments and suggestions.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

I Introduction

The use of grades as a student incentive mechanism as well as a signaling device in elementary through higher education is pervasive. For teachers grades can be seen as a low-financial-cost instrument to induce student effort and for employers and post-secondary institutions as a quantifiable, though imperfect, metric of achievement. Grades are analogous to performance based compensation in agent/employer relationships from which empirical evidence suggests a strong link between compensation and effort.[1]

Theoretical work by Betts (1998) and Costrell (1994) model future labor market earnings as an endogenous function of the achievement groups students sort themselves into based on a grading standard. If a student meets the standard the student signals a higher expected productivity measure than if the student does not meet the standard which is rewarded accordingly in the labor market. Betts goes further to show that with heterogeneous ability levels, a barrier of ability exists such that those with ability above such a barrier will choose to meet the standard and those below will not. Since actual achievement from secondary school goes unobserved by colleges, universities, and employers they must rely in part on the rough measure that comes from grades to determine admittance or wages. Students in turn achieve the necessary amount in school to send the best signal facing their constraint of effort and ability. Then, a naive lever to increase achievement for groups with the requisite ability and excess effort would be to increase the required level of achievement to meet or exceed a standard (i.e. raise grading standards). However, concern must also be given to students at the margin of getting a given grade who may be restricted to one lower or to students that must channel effort from other activities (other than leisure)[2] which could lead to

[1]See, e.g., Ehrenberg and Bognanno (1990); Lazear (2000); Chevalier and Ellison (1999).
[2]Future research investigates this possibility.

potentially worse outcomes for the student and could exacerbate wage inequality among certain groups. It follows that assessing the impact of grading standards and determining an optimal grading standard are important from a policy perspective. Much existing empirical work has evaluated the impact of grading standards while this paper proposes and evaluates a grading standard design.

Recent empirical work by Figlio and Lucas (2004) and Betts and Grogger (2003) presents the costs and benefits to students in terms of degree completion, educational attainment, test score gains, disciplinary infractions, and future labor market earnings from different *school or teacher specific* grading standards and Lillard and DeCicca (2001) looks at the propensity to drop-out of high school as a result of stricter graduation standards. Both Figlio and Lucas and Betts and Grogger find that students exposed to higher grading standards perform better on average than students exposed to lower grading standards for a variety of performance measures and Lillard and DeCicca shows that higher graduation standards can increase dropout rates. Bonesrønning (2004) has also demonstrated that the effects noted by Betts and Grogger and Figlio and Lucas are present for Norwegian high school students that are exposed to a higher grading standard, employing a within school approach similar to Figlio and Lucas. Oettinger (2002) investigates the behavior of students to discrete grade outcomes, finding that students tend to cluster their performance around the visible grade outcomes when faced with an absolute grading standard.

In this paper I introduce and evaluate a grading standard design that allows for multiple grading standards across the grade distribution for any given school. The grading standard design presented in this paper assesses the impact on student outcomes from making it harder for a student to receive a higher grade when at a higher point in the grade distribution than for a student to receive a higher grade when at some point lower in the grade distribution (e.g. harder to receive an A over an A- than to receive a C over a C- in terms of premiums on achievement). This differs from prior empirical research which modeled a single grading standard per school or teacher. The implication from this design is that when grading standards are allowed to be greater at points higher in the grade distribution than points lower test score gains increase on average with the largest gains occurring at the top of the test score distribution. Much of this analysis follows the method used by Betts and Grogger and I also present my replication of their results.

I begin by discussing how grading standards can be modeled at the population level that assumes a random assignment of grading standards. In Section II I describe the High School and Beyond data set, discuss the empirical estimation strategy, and outline the assumptions and limitations concerning the analysis in this paper where non-random assignment is a concern. Results are shown in Section III and Section IV concludes.

A The calculation of grading standards from a population perspective

Previous literature has modeled a single grading standard per school or teacher. In departing from that model, this paper considers a continuum of grading standards at each value of the grade distribution. It is instructive to think of both models from a population perspective and to see the differences therein.

A single grading standard per school[3] can be modeled as the school specific intercept from the following equation:

$$a_{ij} = \alpha_j + \beta g_{ij} + \gamma c_{ij} + \delta X_{ij} + \epsilon_{ij} \tag{1}$$

[3] We can also think of this in terms of teacher grading standards by comparing student outcomes across teachers within a school.

where student i at school j has achievement of a_{ij}, a math GPA of g_{ij} from courses c_{ij}, and ϵ_{ij} is some unobserved idiosyncratic error component with $E_{i \in j}[\epsilon_{ij}] = \bar{\epsilon}$. The vector X includes measures of teacher quality, school quality, and parental resources. Under this model a school's grading standard is given by α_j (Betts and Grogger, 2003). Consider students randomly assigned to schools $1, 2$. If these schools are identical except for the manner in which achievement is awarded with grades conditional upon these factors X then we can express the conditional differences in average achievement at the two schools as:

$$E_{i \in j}[a_{i1} - a_{i2} | X, c_{ij}] = \alpha_1 - \alpha_2 \quad (2)$$

Then, if $\alpha_1 > \alpha_2$ school 1 would be considered to have a higher grading standard than school 2 as it requires more average achievement for the same reward. Since this grading standard is not a function of math GPA, as we move along the grade distribution the standard is being modeled as constant.

Now, consider a different model of grading standards where schools have different standards attached to each level of the GPA distribution. Consider the same setup leading to (1) but with standards at each school varying over the GPA distribution:

$$a_{ij} = \alpha_j + \beta_j g_{ij} + \gamma c_{ij} + \delta X_{ij} + \epsilon_{ij} \quad (3)$$

where the model is identical to (1) except standards include α_j and β_j for each school. Consider students randomly assigned to schools $1, 2$.

$$E_{i \in j}[a_{i1} - a_{i2} | X, c_{ij}, \alpha_1 = \alpha_2] = (\beta_1 - \beta_2) g \quad (4)$$

$$E_{i \in j}[\frac{\partial \Delta^{12} a}{\partial g} | X, c_{ij}, \alpha_1 = \alpha_2] = \beta_1 - \beta_2 \quad (5)$$

It follows that if $\beta_1 > \beta_2$ school 1 has differentially higher standards than school 2 as the average change in achievement necessary to receive a given higher grade at school 1 is greater than that at school 2 with the premium on achievement increasing by a factor of $\beta_1 - \beta_2$ as a student moves along the grade distribution.

It is also natural to question to what extent the growth rate of grading standards across the GPA distribution matters in determining student achievement. The main text of this document describes the effect of differences in the rate of increase of achievement across the GPA distribution as modeled in the population above. In Appendix D, I show results that use proportional changes to construct grading standards rather than differential changes. The results from the two grading standards constructions are very similar.

II Data and methods

A High School and Beyond Survey

I use data from 1980, 1982 and 1984 on the 1980 sophomore cohort from the High School and Beyond (HSB) longitudinal survey. The HSB survey is meant to be representative of the approximately 3.8 million United States students that were sophomores in 1980 at 24,725 high schools. Data were collected from approximately 30,000 students at approximately 1,122 schools in 1980 and 1982, and from probability samples of approximately 15,000 students in 1984. Schools that were identified as having 30% or more of a certain population such as black or Hispanic were oversampled making this analysis particularly useful for looking at the effect of grading standards on those groups.[4]

I focus my analysis on the universe of public school students from the 15,000 student probability sample, a subsample of 11,436 students and 854 schools. I also exclude individuals that were not chosen to participate in the 1980 survey leaving a sample of 10,548 and 854 schools. When constructing the grading standards

[4]Unweighted results are shown. Weighted results are similar.

I require that the student remain at the same school from 10th to 12th grade and that transcripts and a 12th grade test score be available from at least five students at each school. The restriction of the sample to those with five or more representative students was chosen with consideration of the trade-off between appropriately estimating a slope and intercept for each school in addition to having enough power for estimation. Less than two percent of the otherwise eligible students and 10% of the otherwise eligible schools were excluded on this basis. The results are not sensitive to the cutoff of five. This leaves 7,215 students at 678 schools used to construct grading standards with 8,940 students available for the outcome regressions.[5]

Students took exams in math, reading, and science when they were sophomores in 1980 and again in 1982. For the math exam students were administered a fifteen minute exam from which a formula score was computed.[6] One advantage of using the High School and Beyond Survey is that the achievement test administered by NCES was not used to determine teacher or student outcomes. Should students have also been subject to a state test-based accountability system, that assignment would be random. This removes the complication that comes from using No Child Left Behind achievement test populations. In studies that use these student populations an achievement test might alter teacher behavior and so lower the resources available to students far from the pass/fail margin (Neal and Schanzenbach, 2010). Care then would need to be given to the possibility that the achievement distribution affects the probability of neglect. This study is free of that concern.

[5]1,725 (8,940 - 7,215) additional students are available for outcome regressions as they were attending schools in 10th grade that had enough other eligible students to enable the construction of a grading standard for that school.

[6]The formula score awards 1 point for correct answers and subtracts one quarter point for incorrect answers.

Additional parent and school surveys and administrative data complement the student surveys. A transcript survey was conducted for a probability sample of the 1980 sophomores available for the first follow-up. From this survey I am able to construct a math grade point average and variables that detail the type of math course and the quantity for each student.[7] School level surveys were completed by school principals in 1980 from which I construct some of the control variables.

B Estimation models

In estimating grading standards I simply add an interaction term to the original model proposed by Betts and Grogger and will refer to this model as the "Slope and intercept model". Specifically, the model is[8]:

$$A_{ij} = \sum_{j=1}^{n^{SCH}} SCHOOL_{ij}\alpha_j + \sum_{j=1}^{n^{SCH}} (SCHOOL_{ij} * GPA_{ij})\xi_j + \sum_{m=1}^{n^{COR}} \rho_{ijm}\beta_m + GPA_{ij}\gamma + \epsilon_{ij} \quad (6)$$

where A_{ij} is the grade 12 math test score for person i in school j. $SCHOOL_{ij}$ is a dummy variable equal to 1 if student i is in school j and zero otherwise. GPA_{ij} is a calculated math GPA for student i in school j. ρ_{ijm} is the number of courses of type m taken by student i at school j. The grading standard is given by the estimated coefficient on $SCHOOL$ and the coefficient on $(SCHOOL*GPA)$ in equation (6). Differences in scale between math GPA and the standardized math test are given in part by γ and ξ_j for all schools j. For comparison purposes I also present what I refer to as the "Intercept only model" which is the specification from Betts and Grogger and is identical to the model above without the interaction term. The

[7] For details on this construction see Appendix A.

[8] See Appendix A for details on how some of the ambiguous components of this regression are computed.

resulting grading standard in the intercept only model is given by the coefficient on $SCHOOL_{ij}$ from that model.[9]

For the outcome equations I let the effects of the school specific slope and the intercept vary independently.[10] Specifically, I estimate:

$$\begin{aligned} y_{ij} &= \delta_1 \hat{\alpha_j} + \delta_2 \hat{\xi_j} + \tau_1 (10^{th} \text{ grade test score})_{ij} \\ &+ \tau_2 (10^{th} \text{ grade school average test score})_j + X_{ij}\phi + Z_j\theta + \mu_{ij} \end{aligned} \tag{7}$$

Outcomes are 12th-grade test scores and dummy variables for high school graduation and college attendance. Both high school graduation and college attendance are only observed until 1984 which corresponds to the second year of college for a typical college matriculated student in this sample. X represents a vector of characteristics relating to the student, Z a vector of school characteristics, and the effects of grading standards are given by the coefficients δ_1 and δ_2 with the differential effect given by δ_2.[11]

[9] In Betts and Grogger the model that is estimated is:

$$A_{ij} = \sum_{j=1}^{n^{SCH}} SCHOOL_{ij}\alpha_j + \sum_{m=1}^{n^{COR}} \rho_{ijm}\beta_m + GPA_{ij}\gamma + \epsilon_{ij}$$

[10] It is necessary to fix the intercept when looking at the effect of the slope. Without fixing the intercept the slope coefficient would not necessarily be picking up the effect of differentially higher grading standards. To see this, consider school A with an estimated intercept, $\hat{\alpha_A}$, of 14 and an estimated slope, $\hat{\xi_A}$, of 2. Now, consider a school B with a smaller intercept but a bigger slope, in this example let $\hat{\alpha_B} = 2$ and $\hat{\xi_B} = 8$. Then, on the math GPA range from $[0, 2)$ which is an average GPA of an "F" to a "C", school A has higher grading standards as the average achievement necessary from a student at school A to get the same GPA as a student at school B is higher than that from a student at school B. However, this relationship reverses from a GPA between a "C" and an "A" with a student at school B with an average GPA in this range facing higher grading standards than a student at school A. If this example were to occur in the data and δ_1 were left out of equation (7) then δ_2 would be picking up the mixed effect of having higher and lower standards across the grade distribution.

[11] To show the stability of the coefficients across covariate, X and Z vary in different specifications. For details behind each specification see Appendix B.

C The assumptions concerning the analysis in this paper

With the absence of a natural experiment or a true measure of student achievement this study relies on a number of assumptions to identify the effect of grading standards on student outcomes. The most basic assumption is that the 12^{th} grade standardized math test is a meaningful and objective measure of math achievement. In Betts and Grogger they conclude that the test does not appear to be racially biased. I perform a similar test for a gender biased exam and my results are consistent in that specification as well as the one shown.

Another strong assumption is that math courses of a similar name are similar across schools. In comparing a resource rich school to a deprived school we can expect that this will not be true with courses being of higher quality at the resource rich school. The implication is that the higher quality school may incorrectly be assigned a higher grading standard. To reduce the impact this would have on outcomes I include a number of school quality controls in the outcome equation.

When evaluating the impact of grading standards it would be ideal to have student achievement measures prior to any exposure to the grading standards. Unfortunately I do not see students before they enter high school - the first observation occurring in the student's sophomore year. It is possible that some of the impact of grading standards occurs in the first year of high school. To the extent the impact of grading standards is uniformly positive this will result in my estimates understating the true effect of grading standards.

Furthermore, the analysis requires that my controls for school achievement are sufficient to avoid the possibility that the distribution of student achievement causes grading standards. I have found strong evidence that schools do to a large extent tailor grading standards to the within school achievement distribution. To account for this tendency to grade on a curve I include a control for the mean of the school 10^{th} grade test score distribution.[12]

Finally, as I am unable to see the same student under two different and random grading standard regimes it is necessary to make some assumptions about how a student's grading standards are determined. It is possible that students who would have been adversely impacted by high grading standards knew this before enrolling in a public high school and so enrolled in private school or a different public high school. This shifting of students could occur in both directions with those best impacted by high grading standards shifting to those type schools and those best impacted by low standards also shifting. The opportunity to transfer schools for the benefit of the student will be more possible for the resource rich student and since high grading standards are correlated with family income this net bias is likely to inflate the effect we would unconditionally observe.

To try to account for this potential bias I include an extensive list of family controls and lagged student achievement. For identification I assume that the 10^{th} grade math test score serves as a good proxy for prior math achievement before entering high school and that after conditioning on family covariates, prior math achievement is all that affects the student's eventual placement at a particular school.

D How missing observations were handled

Some of the key variables have missing information and to the extent that the missing information is not random my results may be biased. Transcript information is missing for 664 students that are attached to a constructed grading standard.

[12] I also run specifications that include a measure of skewness, an indicator for negative skewness, and a count for the number of students within a school with a 10^{th} grade math score above the 75^{th} percentile of such scores. For all specifications, the results are similar to those shown in this paper.

This missingness is positively correlated with grading standards and a comparison of means shows that these missing students have lower average test scores and lower test score gains. Since they are included in the second stage regression this could result in a downward bias of the effect of grading standards on achievement if their presence in the construction of grading standards would have reduced those variables. To the extent that is possible I also compute regressions where these students are included in the equation constructing grading standards and are assigned a math GPA within one standard deviation above their 12th grade test score which will mechanically lower their school's grading standard. Results are similar.

The problem of a missing 12th grade test score for 1,180 students in my sample is handled differently. Since it is difficult to find a valid instrument for a selection equation in the HSB data I follow Betts and Grogger in first determining the student's percentile rank in the 10th grade test distribution. I then assign the student a 12th grade test score within decile from the 12th grade test score distribution. This method requires the student have a 10th grade test score from which 351 students do not.[13] These students with imputed 12th grade test scores are excluded from the calculation of grading standards but included in outcome regressions.

Another source of missing observations involves students that transfer schools. When looking at outcomes it is plausible that the grading standards caused the 538 students that transferred schools to do so because of poor performance at the original school. These students are linked to the school they attended before the transfer. In what I believe to be a conservative approach I first determine the student's percentile rank in the 10th grade test score distribution. I then assign the student a 12th grade test score within decile from the 12th grade test score

[13] Following the assumption that missingness is random I have also run models where I assign any student missing a 12th grade test score the mean 12th grade grade test score which allows for including the entire 1,180 students with missing 12th grade test scores and my results are similar.

distribution. This method assumes that if these students had not transferred they would have had no improvement in the 12th grade test score distribution.[14]

III Results

Using the fixed effects and the fixed effects interactions as grading standards in a second stage equation presents a number of econometric issues. First, since the grading standards are estimates they are subject to sampling variation that will enter the second stage. Standard errors that do not account for this variation will be smaller than corrected standard errors if the disturbances are uncorrelated but with correlation the direction is ambiguous (Murphy and Topel, 2002).

Second, the estimation method requires that certain estimators in the second stage not be included in the first stage. Therefore, covariates in the second stage may be correlated with the residuals from the first stage which will enter through the grading standards variables (Cardell and Hopkins, 1977).

Furthermore, asymptotic results are based on limits as the number of students gets large. However, as we increase the number of students we must undoubtedly also increase the number of schools. Then, it becomes difficult to form a consistent estimate of the school coefficients as their estimate will depend on still unknown other parameters (Neyman and Scott, 1948).

All standard errors and statistical tests are given with assumptions that the sampling distribution constructed from sampling the data provides a good approximation to the true sampling variation. With an additional assumption that the

[14]The imputation methods resulted in little change over simply dropping students that transferred high schools or that had a missing 12th grade test score. Results with imputed test scores show smaller (though not statistically different) effect sizes for the differential and absolute coefficients from the slope and intercept model (which will be described in Section B). Results from dropping either those with a missing 12th grade test score or those that transferred schools can be seen in Table 18 on page 47 for 12th grade test score as an outcome and in Table 19 on page 48 for the outcomes of high school graduation and college attendance.

student unobservables are correlated within schools I present bootstrapped standard errors with respect to school clusters of 1,000 repetitions from regressions estimated using OLS.

A Outcome: 12^{th} grade test scores

Overall

Table 1 on page 15 displays estimates of the effect of grading standards on 12^{th} grade test scores. Panel A presents my replication of the model from Betts and Grogger and panel B displays results from the slope and intercept model. In both models, the grading standard constructed from school-specific intercepts is labeled as the "absolute" effect and the grading standard constructed from the slope as the "differential" effect. All variables shown including the outcome variable, 12^{th} grade test scores, have been centered at zero with standard deviation one.

Columns (1) through (5) show how the grading standard estimates are stable in response to an increasingly saturated regression model. It is necessary to condition on either the student's 10^{th} grade test score or the student's school's mean 10^{th} grade test score since achievement distributions cause grading practices to some extent. In all regressions shown I include both.

By comparing across columns you will notice that once I condition on the lags the coefficients appear orthogonal to school level controls which include the percent of student's that drop out and an indicator for urban; threats to learning controls which contain indicators for the severity of problems such as student's cutting class and robbery and theft on campus; student demographic controls which include occupational dummies and family income; and to census division dummies. χ^2 statistics are reported from a Wald test that uses the inverse of the cluster, bootstrap covariance matrix when testing whether the additional covariates shown in the column are significantly different than zero.[15] For all groups of controls except the census division dummies the additional controls are shown to matter for both models in determining 12^{th} grade test score.

Looking first at Panel A which displays the replication of the Betts and Grogger results my estimated coefficients are slightly larger than those reported in Betts and Grogger with a one standard deviation increase in absolute grading standards resulting in a 0.1030 standard deviation increase in 12^{th} grade test scores (compared to 0.068 as shown in their paper). I am unable to explain this difference. However, since my replication follows only the guide in the text it is possible that we differed in our measures of the 12^{th} grade test score, imputation methods, and course variable constructions. Since both paper's results depend on a conditional independence assumption these differences in method could cause our results to differ due to insufficient controls or measurement error in either paper.

Turning now to Panel B, column (5) the estimated absolute effect of grading standards is larger than in the intercept only model. The differential effect is also strong and significant. Standard errors are reported with the bootstrap method sampling on school clusters. Bootstrapping the standard errors results in only modest increases in the standard errors of the slope and intercept model as compared to a cluster calculation (less than 5% change for the results shown in column 5); however, the effect of bootstrapping is more substantial in the intercept only model which present standard errors approximately 70% smaller than those when clustering alone is used.

Provided the controls are sufficient for validity of an independence of treatment assumption Table 1 shows that differentially higher grading standards result in an increase in average achievement. A school that has differentially higher grading

[15]This statistic is estimated using the variance weighted distance of a Wald but it lacks a definable likelihood since unobservables are correlated within schools.

standards of one standard deviation above the mean can be expected to have higher achievement by 0.13 standard deviations than a school with average differential grading standards.

Since the effect of grading standards will be driven by preferences, effort, and ability; it is unlikely that the effect will be uniform across groups. The next few sections present results for students at different quantiles of the achievement distribution, also for students of different race and ethnicity, and for men and women separately.

Table 1: Estimated effects of grading standards on 12th grade test scores[a b]

Grading Standard	(1)	(2)	(3)	(4)	(5)
A. Intercept only model					
Absolute effect	0·1079	0·1043	0·1058	0·1018	0·1030
	(0·0095)	(0·0092)	(0·0033)	(0·0091)	(0·0027)
R^2	0·67	0·67	0·67	0·70	0·70
Incremental[c] χ^2		42·**	133·***	900·***	94·***
		[20]	[18]	[82]	[9]
B. Slope and intercept model					
Absolute effect	0·1716	0·1631	0·1688	0·1616	0·1621
	(0·0197)	(0·0183)	(0·0056)	(0·0174)	(0·0176)
Differential effect	0·1366	0·1291	0·1366	0·1280	0·1281
	(0·0182)	(0·0169)	(0·0052)	(0·0160)	(0·0162)
R^2	0·67	0·67	0·67	0·70	0·70
Incremental χ^2		40·**	78·***	979·***	8·
		[20]	[18]	[82]	[9]
Other controls (all regressions in column):					
Lags	X	X	X	X	X
School variables[d]		X	X	X	X
Threats to learning			X	X	X
Student demographics				X	X
Census division dummies					X

[a] Figures in parentheses are standard errors. Sample size is 8589.

[b] All regressions include the student's 10^{th} -grade test score, the student's school's mean 10^{th} -grade test score, and a dummy indicating whether those variables are missing.

[c] χ^2 statistics shown with degrees of freedom in brackets from a Wald test that the additional covariates are jointly no different than zero († p<0.10, * p<0.05, ** p<0.01, *** p<0.001).

[d] Includes dummy variables for fraction of students black, fraction of students who drop out, fraction of class taking remedial math, and miles between high school and four year college/university when missing.

By quantile

I estimate quantile regressions with the same covariates used in column (5) of Table 1 on the previous page at the 10^{th} through the 90^{th} quantiles of the 12^{th} grade test score distribution at increments of $1/20^{th}$.[16] These results are shown graphically in Figure 1 and again in Table 8 on page 34.

Figure 1 presents the grading standard coefficient estimates from the intercept only model and the slope and intercept model with two standard error bands. Focusing on the differential effect from the slope and intercept model we see that the effect is strongest at the upper quantiles with the impact of the effect rising linearly until around the 60^{th} quantile, dipping to the 75^{th} and then ascending more rapidly through the 90^{th} . The effect is largest at the 90^{th} quantile with predicted student gains of 95% more than students in the 10^{th} quantile from a one standard-deviation increase in differential grading standards. The absolute effect from both models has similar patterns with the effect size at the lowest quantile being nearly half that of the most upper. The median from each grading standard is considerably lower than its corresponding mean indicating that the conditional density is asymmetric or that outliers are partially contributing to the effect.[17]

[16]Standard errors are computed using the Koenker and Bassett method in this version of the paper. Since I believe the assumption of a homoscedastic error distribution is violated this should be viewed as an imperfect approximation. Bootstrap standard errors will substitute at some later date.

[17]See Koenker and Hallock (2001) for this discussion.

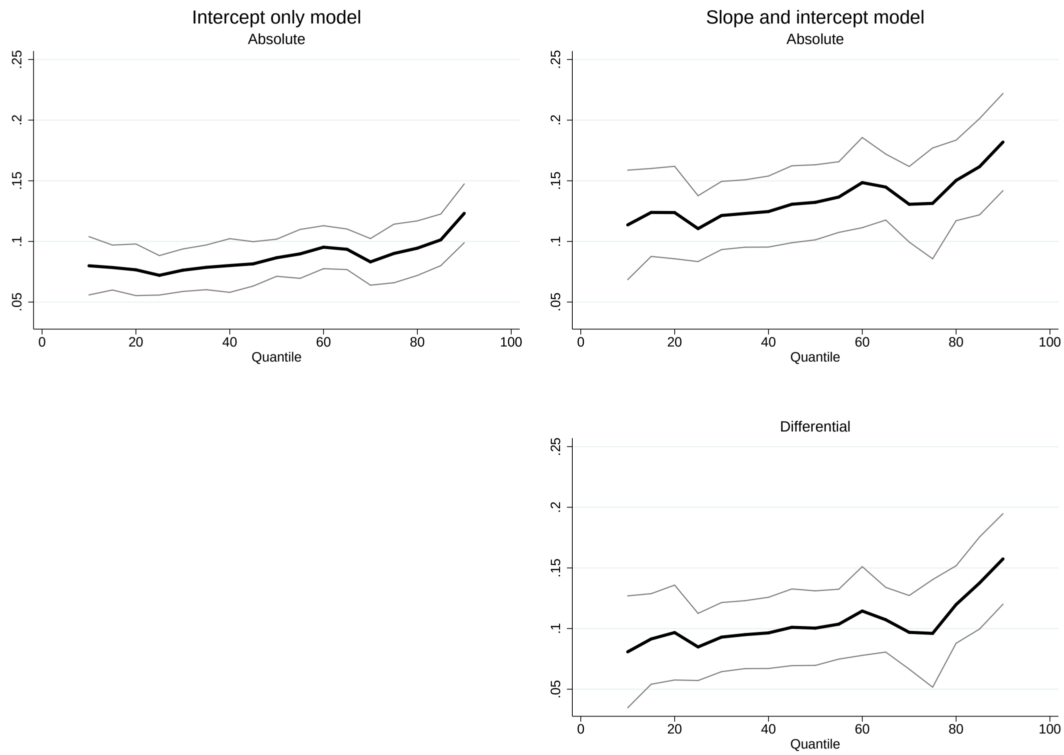

Figure 1: Coefficient estimates from quantile regressions ranging from the 10th to the 90th quantiles in steps of 1/20th with two standard error bands

By gender

A number of studies have concluded that women respond differently to competition with any benefit received from competition being less than that for men.[18] An increase in absolute or differential grading standards could result in more peer-group competition should grading standards result in fewer desirable grade opportunities. In Table 2 on page 19 I split the sample by gender with female results appearing in panels I and male results appearing in panels II. Student demographic variables no longer contain the student's gender. Models were also considered that included the fraction of females per school and results were similar to those shown with the fraction shown to not matter.

[18]See, e.g., Gneezy et al. (2003); Gneezy and Rustichini (2004); Niederle and Vesterlund (2007); and Price (2008).

In each model and for both grading standards considered the effects are largest for males; however, the difference is only statistically significant for the intercept only model given the more imprecise estimates that come from the intercept and slope model. Looking at the column (5) specification from the intercept only model, men stand to gain 40% more than women from a one standard deviation increase in absolute grading standards. The estimated gains for men relative to women from the slope and intercept model is 26% and 21% for the absolute and the differential effects, respectively.

Grading standard parameter estimates by each quantile of the 12^{th} grade test score distribution are presented separately for women and men in Figure 2 and again in Table 9 on page 35 and Table 10 on page 36. The effect is gradually increasing across the achievement distribution for women with a slight but insignificant dip and gain occurring at the 65^{th} percentile of the female test score distribution. For men the effect is fairly constant across quantile until around the 70^{th} quantile where the effect size rises sharply for men with a relative change from the 70^{th} to the 90^{th} quantile of 58%, 66%, and 107% for the absolute effect in the intercept only model, and the absolute, differential effects from the slope and intercept model, respectively. This large gain in the upper tail results in 90/10 ratios of 1.53, 1.86, and 2.39 for the absolute effect in the intercept only model and the absolute and differential effects from the slope and intercept model, respectively. However, the median compared to the bottom quantile presents a very different story with 50/10 ratios of .89, .98, and 1.11, respectively.

Table 2: Estimated effects of grading standards on 12th grade test scores by gender[ab]

Grading Standard	(1)	(2)	(3)	(4)	(5)
A. Intercept only model					
I. Female					
Absolute effect	0·0961	0·0928	0·0948	0·0837	0·0854
	(0·0123)	(0·0126)	(0·0042)	(0·0123)	(0·0050)
II. Male					
Absolute effect	0·1190	0·1153	0·1166	0·1191	0·1197
	(0·0125)	(0·0119)	(0·0042)	(0·0119)	(0·0042)
B. Slope and intercept model					
I. Female					
Absolute effect	0·1603	0·1533	0·1586	0·1412	0·1429
	(0·0230)	(0·0222)	(0·0227)	(0·0221)	(0·0222)
Differential effect	0·1306	0·1258	0·1337	0·1143	0·1152
	(0·0214)	(0·0215)	(0·0211)	(0·0209)	(0·0205)
II. Male					
Absolute effect	0·1809	0·1708	0·1774	0·1796	0·1794
	(0·0242)	(0·0241)	(0·0091)	(0·0228)	(0·0235)
Differential effect	0·1417	0·1323	0·1391	0·1399	0·1395
	(0·0230)	(0·0225)	(0·0089)	(0·0218)	(0·0227)
Other controls (all regressions in column):					
Lags	X	X	X	X	X
School variables[c]		X	X	X	X
Threats to learning			X	X	X
Student demographics				X	X
Census division dummies					X

[a]Figures in parentheses are standard errors. Sample size is 4277 for panels I and 4312 for panels II.

[b]All regressions include the student's 10^{th} -grade test score, the student's school's mean 10^{th} -grade test score, and a dummy indicating whether those variables are missing.

[c]Includes dummy variables for fraction of students black, fraction of students who drop out, fraction of class taking remedial math, and miles between high school and four year college/university when missing.

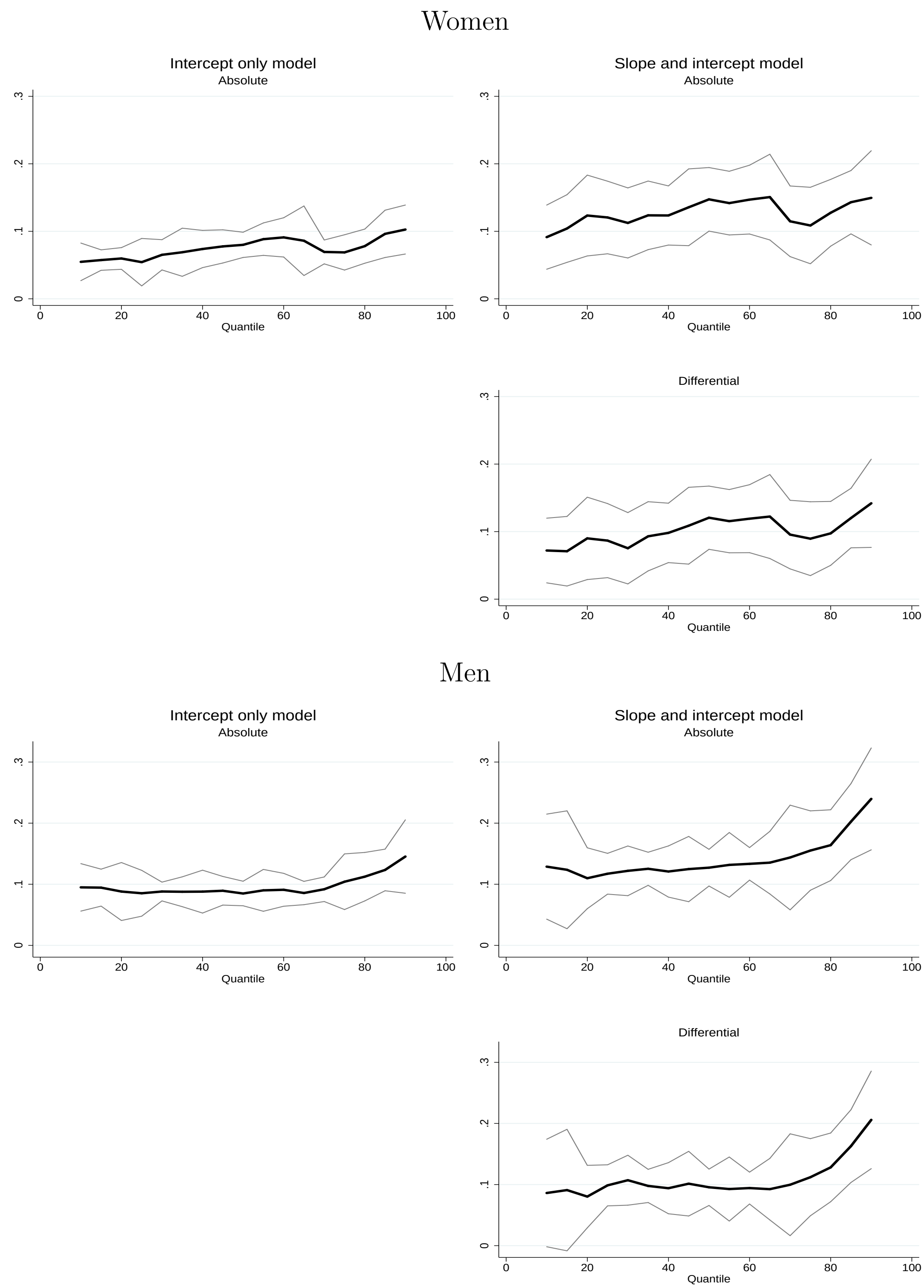

Figure 2: Coefficient estimates from quantile regressions ranging from the 10^{th} to the 90^{th} quantiles in steps of $1/20^{th}$ with two standard error bands by gender

By race/ethnicity

I present results for Hispanic, black, and white separately in Table 3 on the following page for the intercept only model shown in panel A and the slope and intercept model shown in panel B. In both models the effect is strongest for blacks with the smallest effect being shared by whites and Hispanics. Student demographics variables exclude race/ethnicity.[19]

Focusing on the differential effect from column (5) Table 3 on the next page shows that black students have predicted gains in average test scores of 1/5 of a standard deviation from a one standard deviation change in grading standards which is approximately twice the effect for Hispanic students and a little more than one and a half that of white students. The absolute grading standard for blacks relative to Hispanics varies across the models being 38% larger in the intercept only model and 68% larger in the slope and intercept model. The relative difference is similar for both models when comparing black students to white students with black students gaining approximately 60% more than white students in both specifications.

[19]The effect for blacks is in stark contrast to the effect noted by Betts and Grogger where it was noted that the effect was smallest for blacks in comparison to whites and Hispanics. This difference in effect is present when I restrict the sample to those without imputed test scores, to those that did not transfer high schools, and when I restrict the analysis to the core covariates used in the Betts and Grogger paper of lags, school variables, and student demographics.

Table 3: Estimated effects of grading standards on 12th grade test scores by race[ab]

Grading Standard	(1)	(2)	(3)	(4)	(5)
A. Intercept only model					
I. Hispanic only					
Absolute effect	0·1183	0·1157	0·1200	0·1052	0·1151
	(0·0183)	(0·0115)	(0·0195)	(0·0040)	(0·0030)
II. Black only					
Absolute effect	0·1179	0·1482	0·1579	0·1490	0·1594
	(0·0286)	(0·0291)	(0·0114)	(0·0248)	(0·0309)
III. White only					
Absolute effect	0·0956	0·0912	0·0923	0·0994	0·0996
	(0·0114)	(0·0115)	(0·0038)	(0·0113)	(0·0118)
B. Slope and intercept model					
I. Hispanic only					
Absolute effect	0·1689	0·1675	0·1772	0·1511	0·1546
	(0·0336)	(0·0344)	(0·0204)	(0·0205)	(0·0216)
Differential effect	0·1111	0·1115	0·1281	0·1032	0·1050
	(0·0306)	(0·0294)	(0·0178)	(0·0182)	(0·0192)
II. Black only					
Absolute effect	0·2136	0·2519	0·2658	0·2535	0·2607
	(0·0455)	(0·0476)	(0·0417)	(0·0479)	(0·0477)
Differential effect	0·1655	0·1870	0·2073	0·1903	0·2045
	(0·0424)	(0·0434)	(0·0394)	(0·0456)	(0·0430)
III. White only					
Absolute effect	0·1575	0·1469	0·1513	0·1604	0·1615
	(0·0199)	(0·0206)	(0·0089)	(0·0062)	(0·0204)
Differential effect	0·1298	0·1201	0·1241	0·1300	0·1312
	(0·0191)	(0·0197)	(0·0086)	(0·0063)	(0·0197)
Other controls (all regressions in column):					
Lags	X	X	X	X	X
School variables[c]		X	X	X	X
Threats to learning			X	X	X
Student demographics				X	X
Census division dummies					X

[a]Figures in parentheses are standard errors. Sample size is 1809 for panels I, 1061 for panels II, and 5244 for panels III.

[b]All regressions include the student's 10^{th} -grade test score, the student's school's mean 10^{th} -grade test score, and a dummy indicating whether those variables are missing.

[c]Includes dummy variables for fraction of students black, fraction of students who drop out, fraction of class taking remedial math, and miles between high school and four year college/university when missing.

B Outcome: High school graduation and college attendance by 1984

Overall

In Table 4 on the following page I look at the effect of grading standards on high school graduation and in Table 5 on page 25 I look at the effect of grading standards on college attendance at a four year college or university with linear probability models. A student is calculated as having graduated high school if by 1984 the student has received a diploma and not a GED. A student is calculated as having attended college if by 1984 the student has earned some positive amount of credit hours at a four year college or university.

The effect is insignificant for both outcomes, both models and grading standards considered. Betts and Grogger have similar findings and suggest that this may be a result of the largest effect of grading standards occurring at upper quantiles of the test score distribution where students are more likely to graduate high school and go to college anyway. Despite modest gains at the lower quantiles of the test score distribution the gains may not be enough to induce a change in preferences or a relative decrease in outside options which could be why these quantiles do not contribute to an overall positive effect. It could also be that the effect is mixed for these groups with some students becoming discouraged by higher grading standards and dropping out or not attending college while causing others to earn a degree or attend college with the effects on these different groups netting out. Of course, these results are conditional on education decisions being made by 1984 so I cannot identify any effect of grading standards that might postpone education decisions beyond 1984.

Table 4: Estimated effects of grading standards on high school graduation[ab]

Grading Standard	(1)	(2)	(3)	(4)	(5)
A. Intercept only model					
Absolute effect	−0·0013	−0·0006	0·0004	0·0040	0·0037
	(0·0052)	(0·0053)	(0·0017)	(0·0044)	(0·0017)
R^2	0·06	0·07	0·07	0·26	0·27
Incremental[c] χ^2		43·**	290·***	1575·***	105·***
		[20]	[18]	[82]	[9]
B. Slope and intercept model					
Absolute effect	0·0042	0·0039	0·0079	0·0107	0·0107
	(0·0098)	(0·0100)	(0·0031)	(0·0082)	(0·0082)
Differential effect	0·0104	0·0107	0·0158	0·0134	0·0126
	(0·0106)	(0·0110)	(0·0036)	(0·0093)	(0·0091)
R^2	0·06	0·07	0·07	0·26	0·27
Incremental χ^2		41·**	274·***	1447·***	20·*
		[20]	[18]	[82]	[9]
Other controls (all regressions in column):					
Lags	X	X	X	X	X
School variables[d]		X	X	X	X
Threats to learning			X	X	X
Student demographics				X	X
Census division dummies					X

[a]Figures in parentheses are standard errors. Sample size is 8679.

[b]All regressions include the student's 10^{th} -grade test score, the student's school's mean 10^{th} -grade test score, and a dummy indicating whether those variables are missing.

[c]χ^2 statistics shown with degrees of freedom in brackets from a Wald test that the additional covariates are jointly no different than zero († p<0.10, * p<0.05, ** p<0.01, *** p<0.001).

[d]Includes dummy variables for fraction of students black, fraction of students who drop out, fraction of class taking remedial math, and miles between high school and four year college/university when missing.

Table 5: Estimated effects of grading standards on college attendance by 1984[a b]

Grading Standard	(1)	(2)	(3)	(4)	(5)
A. Intercept only model					
Absolute effect	−0·0122	−0·0091	−0·0073	−0·0065	−0·0016
	(0·0087)	(0·0078)	(0·0028)	(0·0072)	(0·0030)
R^2	0·16	0·17	0·18	0·26	0·27
Incremental[c] χ^2		66·***	327·***	1353·***	528·***
		[20]	[18]	[82]	[9]
B. Slope and intercept model					
Absolute effect	−0·0215	−0·0201	−·0160	−0·0141	−0·0075
	(0·0134)	(0·0135)	(0·0044)	(0·0121)	(0·0119)
Differential effect	−0·0122	−0·0106	−0·0069	−0·0087	−0·0029
	(0·0129)	(0·0127)	(0·0044)	(0·0117)	(0·0115)
R^2	0·16	0·17	0·18	0·26	0·27
Incremental χ^2		68·***	195·***	1251·***	67·***
		[20]	[18]	[82]	[9]
Other controls (all regressions in column):					
Lags	X	X	X	X	X
School variables[d]		X	X	X	X
Threats to learning			X	X	X
Student demographics				X	X
Census division dummies					X

[a]Figures in parentheses are standard errors. Sample size is 8305.

[b]All regressions include the student's 10th -grade test score, the student's school's mean 10th -grade test score, and a dummy indicating whether those variables are missing.

[c]χ^2 statistics shown with degrees of freedom in brackets from a Wald test that the additional covariates are jointly no different than zero († p<0.10, * p<0.05, ** p<0.01, *** p<0.001).

[d]Includes dummy variables for fraction of students black, fraction of students who drop out, fraction of class taking remedial math, and miles between high school and four year college/university when missing.

By race/ethnicity

In Table 6 on the following page I show estimates of the effect of grading standards on high school graduation by race/ethnicity, again with linear probability models. For each model I do not find an effect of grading standards on graduation. This is in contrast to that observed by Betts and Grogger where they uncovered negative effects for blacks and Hispanics. While the coefficient on the absolute grading standard for the black only models is consistently negative across specification the size of the effect is very small with large standard errors, so large that I cannot with any reasonable uncertainty say that the effect is negative. I also estimate separate models by race/ethnicity for college attendance by 1984 which can be seen in Table 7 on page 28. Similarly, I see no effect which is a conclusion consistent with that reached by Betts and Grogger.

Table 6: Estimated effects of grading standards on high school graduation by race[ab]

Grading Standard	(1)	(2)	(3)	(4)	(5)
A. Intercept only model					
I. Hispanic only					
Absolute effect	0·0127	0·0140	0·0143	0·0082	0·0046
	(0·0115)	(0·0119)	(0·0074)	(0·0072)	(0·0022)
II. Black only					
Absolute effect	−0·0087	−0·0068	−0·0032	−0·0055	−0·0147
	(0·0142)	(0·0156)	(0·0075)	(0·0141)	(0·0190)
III. White only					
Absolute effect	−0·0002	−0·0011	−0·0001	0·0065	0·0068
	(0·0063)	(0·0067)	(0·0029)	(0·0021)	(0·0062)
B. Slope and intercept model					
I. Hispanic only					
Absolute effect	0·0184	0·0190	0·0221	0·0047	−0·0019
	(0·0188)	(0·0195)	(0·0067)	(0·0071)	(0·0062)
Differential effect	0·0117	0·0148	0·0221	−0·0027	−0·0094
	(0·0184)	(0·0183)	(0·0066)	(0·0067)	(0·0061)
II. Black only					
Absolute effect	−0·0208	−0·0156	−0·0048	−0·0150	−0·0257
	(0·0254)	(0·0280)	(0·0239)	(0·0276)	(0·0175)
Differential effect	0·0065	0·0161	0·0249	0·0203	0·0100
	(0·0244)	(0·0246)	(0·0199)	(0·0246)	(0·0161)
III. White only					
Absolute effect	0·0098	0·0059	0·0087	0·0183	0·0188
	(0·0125)	(0·0141)	(0·0136)	(0·0055)	(0·0044)
Differential effect	0·0175	0·0150	0·0186	0·0223	0·0223
	(0·0139)	(0·0159)	(0·0152)	(0·0061)	(0·0050)
Other controls (all regressions in column):					
Lags	X	X	X	X	X
School variables[c]		X	X	X	X
Threats to learning			X	X	X
Student demographics				X	X
Census division dummies					X

[a]Figures in parentheses are standard errors. Sample size is 1836 for panels I, 1095 for panels II, and 5273 for panels III.

[b]All regressions include the student's 10^{th} -grade test score, the student's school's mean 10^{th} -grade test score, and a dummy indicating whether those variables are missing.

[c]Includes dummy variables for fraction of students black, fraction of students who drop out, fraction of class taking remedial math, and miles between high school and four year college/university when missing.

Table 7: Estimated effects of grading standards on college attendance by 1984 by race[a][b]

Grading Standard	(1)	(2)	(3)	(4)	(5)
A. Intercept only model					
I. Hispanic only					
Absolute effect	0·0080	0·0121	0·0157	0·0119	0·0172
	(0·0123)	(0·0115)	(0·0074)	(0·0122)	(0·0027)
II. Black only					
Absolute effect	−0·0337	−0·0250	−0·0196	−0·0269	−0·0111
	(0·0211)	(0·0210)	(0·0160)	(0·0215)	(0·0105)
III. White only					
Absolute effect	−0·0075	−0·0119	−0·0104	−0·0013	0·0031
	(0·0094)	(0·0097)	(0·0036)	(0·0031)	(0·0091)
B. Slope and intercept model					
I. Hispanic only					
Absolute effect	0·0078	0·0054	0·0133	0·0069	0·0103
	(0·0197)	(0·0202)	(0·0067)	(0·0069)	(0·0082)
Differential effect	0·0043	0·0085	0·0159	0·0073	0·0105
	(0·0189)	(0·0177)	(0·0066)	(0·0063)	(0·0077)
II. Black only					
Absolute effect	−0·0420	−0·0226	−0·0156	−0·0300	−0·0117
	(0·0374)	(0·0368)	(0·0084)	(0·0373)	(0·0366)
Differential effect	−0·0458	−0·0388	−0·0248	−0·0466	−0·0229
	(0·0321)	(0·0330)	(0·0081)	(0·0346)	(0·0343)
III. White only					
Absolute effect	−0·0141	−0·0224	−0·0203	−0·0056	0·0008
	(0·0161)	(0·0167)	(0·0168)	(0·0153)	(0·0144)
Differential effect	−0·0008	−0·0084	−0·0066	0·0066	0·0109
	(0·0161)	(0·0167)	(0·0164)	(0·0153)	(0·0143)
Other controls (all regressions in column):					
Lags	X	X	X	X	X
School variables[c]		X	X	X	X
Threats to learning			X	X	X
Student demographics				X	X
Census division dummies					X

[a]Figures in parentheses are standard errors. Sample size is 1746 for panels I, 1024 for panels II, and 5086 for panels III.

[b]All regressions include the student's 10^{th} -grade test score, the student's school's mean 10^{th} -grade test score, and a dummy indicating whether those variables are missing.

[c]Includes dummy variables for fraction of students black, fraction of students who drop out, fraction of class taking remedial math, and miles between high school and four year college/university when missing.

IV Conclusion

This paper has shown that incorporating a continuum of grading standards rather than a single, absolute standard can raise (lower) average achievement when the grading standards increase (decrease) across the math GPA distribution. That is, demanding more achievement from students that wish to go from an A- to an A than students that would like to earn a C rather than a C- can lead to greater average test scores overall. The effect of grading standards on 12^{th} grade test scores is greatest at the top of the achievement distribution, larger for men than women and larger for blacks than whites and Hispanics. No effect from grading standards on high school graduation or college attendance two years post expected high school graduation has been found.

APPENDIX

A Details on grading standard estimation (equation (6))

How math GPA is computed (GPA_{ij})

Full high school transcripts were provided for a probability sample from the sophomore cohort that participated in the 1982 survey. Using the course name I identify math courses and compute a math specific GPA for each student using the school specific coding shown in Figure 3 on the following page after assigning letter grades the appropriate GPA on a 4.33 scale with an F earning a zero, D- a 0.7, and an A+ a 4.33. A fail (Figure 3 #15) is given a value of zero while pass, incomplete, satisfactory, unsatisfactory, withdraw or audit (Figure 3 #14, 16, 17, 18) are not used in either computing the GPA or the count of courses completed by the student. Students are given credit for the highest grade received per course.

To give courses the proper credit toward a GPA, quarter course credits count 1/4 toward a GPA, semester credits 1/2, and trimester 1/3. If a student retakes a course the highest grade counts toward the GPA.

How course activity was computed (ρ_{ijm})

From the transcript data courses were put into categories of similar names such as "Algebra I" or "Introduction to Algebra" and these categories were represented independently in equation (6). If a student fails the course with no successful retake no credit in course count is given to the student. If a student received a grade greater than zero that counted toward a math GPA as described above then the course counted toward the student's total for that category of courses. The ideal would be to compare students with precisely the same courses but that is not possible in this data.

GRADES

Code	If failure is Below 60	If failure is Below 65	If failure is Below 70
1	98-100 A+	98-100 A+	99-100 A+
2	93-97 A	95-97 A	96-98 A
3	90-92 A-	92-94 A-	94-95 A-
4	87-89 B+	89-91 B+	92-93 B+
5	83-86 B	86-88 B	88-91 B
6	80-82 B-	83-85 B-	86-87 B-
7	77-79 C+	80-82 C+	84-85 C+
8	73-76 C	77-79 C	80-83 C
9	70-72 C-	74-76 C-	78-79 C-
10	67-69 D+	71-73 D+	76-77 D+
11	63-66 D	68-70 D	72-75 D
12	60-62 D-	65-67 D-	70-71 D-
13	Below 60 = F	Below 65 = F	Below 70 = F
14	Pass - for a course graded on a pass-fail basis; or Satisfactory - for a course graded on a satisfactory-unsatisfactory basis		
15	Fail - for a course graded on a pass-fail basis; or Unsatisfactory - for a course graded on a satisfactory-unsatisfactory basis		
16	Incomplete		
17	Withdraw		
18	Audit or Registered		

Figure 3: Detail on how course grades are represented in the survey. Reprinted from "High School and Beyond Transcripts Survey", NORC, 1983

B Details on outcome equations (equation (7))

Many of the tables display five specifications that increasingly add covariates to show the sensitivity of grading standards to additional controls. For such tables all specifications shown include lags; four include lags and school variables; three include lags, school variables, and threats to learning; etc. The details of these control variables follow:

Lags:	10^{th} grade standardized test score, Student's school's mean 10^{th} grade standardized test score, and indicator variables for when those values are missing.
School variables:	Indicator for urban, percent of students that are black, percent of students that drop out, miles from a four year college or university, percent of students taking remedial math, and dummy variables representing the 10 quantiles of high school spending per pupil.
Threats to learning:	Dummy variables of "minor or never", "serious or moderate" or whether the principal did not respond to the question "To what degree is each of these matters a problem in your high school?" for the following: 1. student absenteeism 2. students cutting class 3. teacher absenteeism 4. teachers lack commitment or motivation 5. robbery or theft 6. student use of drugs or alcohol

Student demographics:	gender, race dummies, two-parent family indicator and indicator when missing, parent's highest education level dummies, yearly family income in 8^{ths} dummies (top coded at 50K in 1980 $'s), number of siblings dummies, father's occupation dummies, mother's occupation dummies, number of older siblings dummies.
Census division dummies:	Census division dummies.

C Quantile regression estimates

The following three tables list the coefficients and standard errors that are shown visually in Figure 1 on page 17 and Figure 2 on page 20.

Table 8: Quantile regression estimates of the effect of grading standards on 12th grade test scores[ab]

	Intercept only model	Slope and intercept model	
	Absolute	Absolute	Differential
OLS			
Mean	0·1030 (0·0027)	0·1621 (0·0176)	0·1281 (0·0162)
Quantile			
10	0·0799 (0·0120)	0·1136 (0·0226)	0·0808 (0·0231)
15	0·0785 (0·0093)	0·1239 (0·0181)	0·0914 (0·0187)
20	0·0766 (0·0106)	0·1238 (0·0191)	0·0967 (0·0196)
25	0·0721 (0·0081)	0·1106 (0·0136)	0·0848 (0·0138)
30	0·0763 (0·0088)	0·1214 (0·0141)	0·0930 (0·0142)
35	0·0786 (0·0092)	0·1230 (0·0139)	0·0950 (0·0140)
40	0·0801 (0·0111)	0·1246 (0·0146)	0·0964 (0·0147)
45	0·0815 (0·0091)	0·1307 (0·0159)	0·1011 (0·0158)
50	0·0866 (0·0076)	0·1322 (0·0155)	0·1004 (0·0154)
55	0·0897 (0·0101)	0·1366 (0·0146)	0·1036 (0·0144)
60	0·0953 (0·0088)	0·1485 (0·0186)	0·1145 (0·0183)
65	0·0935 (0·0084)	0·1448 (0·0136)	0·1073 (0·0133)
70	0·0832 (0·0096)	0·1306 (0·0156)	0·0969 (0·0152)
75	0·0901 (0·0121)	0·1314 (0·0228)	0·0960 (0·0222)
80	0·0945 (0·0112)	0·1503 (0·0166)	0·1198 (0·0160)
85	0·1013 (0·0106)	0·1616 (0·0198)	0·1376 (0·0190)
90	0·1232 (0·0121)	0·1819 (0·0200)	0·1574 (0·0187)

[a]Figures in parentheses are standard errors.
[b]All regressions include lags, school variables, threats to learning, student demographic, and census division dummy controls. For details on these controls see Appendix B.

Table 9: Quantile regression estimates of the effect of grading standards on 12th grade test scores for women[a][b]

	Intercept only model	Slope and intercept model	
	Absolute	Absolute	Differential
OLS			
Mean	0·0854 (0·0050)	0·1429 (0·0222)	0·1152 (0·0205)
Quantile			
10	0·0548 (0·0139)	0·0914 (0·0238)	0·0720 (0·0240)
15	0·0574 (0·0076)	0·1042 (0·0251)	0·0710 (0·0258)
20	0·0598 (0·0080)	0·1234 (0·0299)	0·0900 (0·0305)
25	0·0543 (0·0176)	0·1205 (0·0269)	0·0867 (0·0274)
30	0·0652 (0·0112)	0·1124 (0·0260)	0·0754 (0·0264)
35	0·0690 (0·0178)	0·1237 (0·0254)	0·0931 (0·0256)
40	0·0738 (0·0138)	0·1235 (0·0219)	0·0981 (0·0220)
45	0·0777 (0·0123)	0·1356 (0·0284)	0·1088 (0·0284)
50	0·0800 (0·0093)	0·1474 (0·0235)	0·1206 (0·0234)
55	0·0884 (0·0120)	0·1418 (0·0236)	0·1155 (0·0234)
60	0·0910 (0·0145)	0·1470 (0·0255)	0·1192 (0·0252)
65	0·0861 (0·0257)	0·1507 (0·0318)	0·1223 (0·0311)
70	0·0695 (0·0088)	0·1148 (0·0261)	0·0956 (0·0254)
75	0·0688 (0·0131)	0·1086 (0·0283)	0·0895 (0·0273)
80	0·0780 (0·0126)	0·1275 (0·0247)	0·0974 (0·0237)
85	0·0963 (0·0175)	0·1431 (0·0234)	0·1201 (0·0220)
90	0·1027 (0·0181)	0·1496 (0·0349)	0·1420 (0·0327)

[a]Figures in parentheses are standard errors.
[b]All regressions include lags, school variables, threats to learning, student demographic, and census division dummy controls. For details on these controls see Appendix B.

Table 10: Quantile regression estimates of the effect of grading standards on 12th grade test scores for men[a][b]

	Intercept only model	Slope and intercept model	
	Absolute	Absolute	Differential
OLS			
Mean	0·1197 (0·0042)	0·1794 (0·0235)	0·1395 (0·0227)
Quantile			
10	0·0949 (0·0194)	0·1288 (0·0430)	0·0862 (0·0440)
15	0·0945 (0·0151)	0·1237 (0·0482)	0·0910 (0·0497)
20	0·0881 (0·0237)	0·1099 (0·0249)	0·0803 (0·0256)
25	0·0853 (0·0188)	0·1172 (0·0167)	0·0987 (0·0168)
30	0·0881 (0·0077)	0·1220 (0·0203)	0·1071 (0·0204)
35	0·0877 (0·0122)	0·1253 (0·0135)	0·0978 (0·0136)
40	0·0880 (0·0175)	0·1209 (0·0209)	0·0940 (0·0209)
45	0·0894 (0·0117)	0·1249 (0·0267)	0·1015 (0·0265)
50	0·0849 (0·0100)	0·1272 (0·0150)	0·0956 (0·0149)
55	0·0900 (0·0171)	0·1317 (0·0265)	0·0927 (0·0262)
60	0·0910 (0·0135)	0·1334 (0·0133)	0·0943 (0·0130)
65	0·0857 (0·0095)	0·1354 (0·0256)	0·0925 (0·0252)
70	0·0918 (0·0101)	0·1438 (0·0428)	0·0997 (0·0417)
75	0·1042 (0·0228)	0·1551 (0·0325)	0·1120 (0·0315)
80	0·1124 (0·0198)	0·1640 (0·0290)	0·1281 (0·0281)
85	0·1234 (0·0170)	0·2025 (0·0311)	0·1629 (0·0298)
90	0·1454 (0·0300)	0·2397 (0·0417)	0·2059 (0·0399)

[a]Figures in parentheses are standard errors.
[b]All regressions include lags, school variables, threats to learning, student demographic, and census division dummy controls. For details on these controls see Appendix B.

D Semilog Results

I also consider a model that compares schools with different growth rates in grading standards across the GPA distribution to determine if grading standards constructed in this manner are similar to those shown in the main body of this paper. The answer is to a large extent yes. To construct grading standard growth rates I run a semilog model in the first stage. Specifically, at the reduced form, the first stage equation at the population is:

$$\ln a_{ij} = \alpha_j + \beta_j g_{ij} + \gamma c_{ij} + \delta X_{ij} + \epsilon_{ij} \tag{8}$$

where the model is identical to equation (3) except I take the natural logarithm of 12^{th} grade achievement. Consider students randomly assigned to schools $1, 2$.

$$E_{i \in j}[\ln a_{i1} - \ln a_{i2} | X, c_{ij}, \alpha_1 = \alpha_2] = (\beta_1 - \beta_2) g \tag{9}$$

Then, if the boundary of the distribution does not depend on the distribution itself and sufficient smoothness conditions exist:

$$E_{i \in j}[\frac{\frac{\partial a_{i1}}{\partial g}}{a_{i1}} - \frac{\frac{\partial a_{i2}}{\partial g}}{a_{i2}} | X, c_{ij}, \alpha_1 = \alpha_2] = \beta_1 - \beta_2 \tag{10}$$

It follows that if $\beta_1 > \beta_2$ school 1 has proportionally higher standards than school 2 as the average growth in achievement necessary to receive a given grade at school 1 is greater than that at school 2 with the premium on achievement growing by a factor of $\beta_1 - \beta_2$.

Results follow in Table 11 on page 39 through Table 17 on page 45. The tables reproduce the results from the intercept only model in Panels A as shown in Tables 1 on page 15 through 7 on page 28 for comparison purposes and show results from this new model which I will refer to as the "Growth and intercept model" in Panels B. Standard errors shown in Panels B of these tables are computed via clustering at the school level without the use of the bootstrap method. As in the body of the text, coefficient estimates have been obtained from different model specifications that become increasingly saturated. The controls used in each regression are listed at the bottom of the table.

Looking first at Table 11, notice that both the absolute effect and the proportional effect are stable across specification. In column (5) we see that a one standard deviation increase in proportional standards results in a 1/10 of a standard deviation increase in average achievement. This is roughly the same effect as estimated in Table 1 on page 15 from differential standards. In Table 12 on page 40 coefficients are estimated separately for men and for women. A similar difference in effect is shown for men and women from this model as that obtained from the slope and intercept model with men shown to have larger gains from higher grading standards than women. Turning now to Table 13 on page 41, the growth and intercept grading standard model estimates different effect relationships by race than those obtained from the slope and intercept model. In the growth and intercept model I estimate effect sizes similar in magnitude for black and white students with a magnitude greater than that for Hispanics. This is in contrast to the results obtained from both the intercept only model and the slope and intercept model where it was found that the effect sizes were similar for whites and Hispanics and smaller in magnitude than that for blacks. The remaining tables, Tables 14 through 17, show no effect of proportional grading standards on high school graduation or college attendance by 1984 which is a conclusion consistent with the results shown in the main body of this paper.

Table 11: Estimated effects of grading standards on 12th grade test scores-semilog model[ab]

Grading Standard	(1)	(2)	(3)	(4)	(5)
A. Intercept only model					
Absolute effect	0·1079	0·1043	0·1058	0·1018	0·1030
	(0·0095)	(0·0092)	(0·0033)	(0·0091)	(0·0027)
R^2	0·67	0·67	0·67	0·70	0·70
Incremental[c] χ^2		42·**	133·***	900·***	94·***
		[20]	[18]	[82]	[9]
B. Growth and intercept model					
Absolute effect	0·1432	0·1308	0·1399	0·1317	0·1301
	(0·0200)	(0·0188)	(0·0193)	(0·0185)	(0·0194)
Proportional effect	0·1046	0·0926	0·1032	0·0926	0·0911
	(0·0192)	(0·0180)	(0·0184)	(0·0182)	(0·0189)
R^2	0·66	0·67	0·67	0·69	0·69
Incremental χ^2		40·**	54·***	773·***	5·
		[20]	[18]	[82]	[9]
Other controls (all regressions in column):					
Lags	X	X	X	X	X
School variables[d]		X	X	X	X
Threats to learning			X	X	X
Student demographics				X	X
Census division dummies					X

[a]Figures in parentheses are standard errors. Sample size is 8589.

[b]All regressions include the student's 10^{th} -grade test score, the student's school's mean 10^{th} -grade test score, and a dummy indicating whether those variables are missing.

[c]χ^2 statistics shown with degrees of freedom in brackets from a Wald test that the additional covariates are jointly no different than zero († p<0.10, * p<0.05, ** p<0.01, *** p<0.001).

[d]Includes dummy variables for fraction of students black, fraction of students who drop out, fraction of class taking remedial math, and miles between high school and four year college/university when missing.

Table 12: Estimated effects of grading standards on 12th grade test scores by gender-semilog model[a][b]

Grading Standard	(1)	(2)	(3)	(4)	(5)
A. Intercept only model					
I. Female					
Absolute effect	0·0961	0·0928	0·0948	0·0837	0·0854
	(0·0123)	(0·0126)	(0·0042)	(0·0123)	(0·0050)
II. Male					
Absolute effect	0·1190	0·1153	0·1166	0·1191	0·1197
	(0·0125)	(0·0119)	(0·0042)	(0·0119)	(0·0042)
B. Growth and intercept model					
I. Female					
Absolute effect	0·1304	0·1161	0·1230	0·1042	0·1024
	(0·0226)	(0·0221)	(0·0226)	(0·0209)	(0·0217)
Proportional effect	0·1002	0·0880	0·0979	0·0786	0·0763
	(0·0212)	(0·0211)	(0·0217)	(0·0201)	(0·0208)
II. Male					
Absolute effect	0·1545	0·1433	0·1549	0·1558	0·1552
	(0·0269)	(0·0253)	(0·0256)	(0·0253)	(0·0267)
Proportional effect	0·1091	0·0975	0·1084	0·1049	0·1049
	(0·0253)	(0·0236)	(0·0238)	(0·0245)	(0·0258)
Other controls (all regressions in column):					
Lags	X	X	X	X	X
School variables[c]		X	X	X	X
Threats to learning			X	X	X
Student demographics				X	X
Census division dummies					X

[a]Figures in parentheses are standard errors. Sample size is 4277 for panels I and 4312 for panels II.

[b]All regressions include the student's 10^{th} -grade test score, the student's school's mean 10^{th} -grade test score, and a dummy indicating whether those variables are missing.

[c]Includes dummy variables for fraction of students black, fraction of students who drop out, fraction of class taking remedial math, and miles between high school and four year college/university when missing.

Table 13: Estimated effects of grading standards on 12th grade test scores by race-semilog model[a b]

Grading Standard	(1)	(2)	(3)	(4)	(5)
A. Intercept only model					
I. Hispanic only					
Absolute effect	0·1183	0·1157	0·1200	0·1052	0·1151
	(0·0183)	(0·0115)	(0·0195)	(0·0040)	(0·0030)
II. Black only					
Absolute effect	0·1179	0·1482	0·1579	0·1490	0·1594
	(0·0286)	(0·0291)	(0·0114)	(0·0248)	(0·0309)
III. White only					
Absolute effect	0·0956	0·0912	0·0923	0·0994	0·0996
	(0·0114)	(0·0115)	(0·0038)	(0·0113)	(0·0118)
B. Growth and intercept model					
I. Hispanic only					
Absolute effect	0·1306	0·1230	0·1316	0·1076	0·1060
	(0·0358)	(0·0351)	(0·0351)	(0·0339)	(0·0353)
Proportional effect	0·0859	0·0829	0·0960	0·0667	0·0652
	(0·0295)	(0·0275)	(0·0276)	(0·0273)	(0·0280)
II. Black only					
Absolute effect	0·1696	0·1734	0·1929	0·1634	0·1546
	(0·0360)	(0·0406)	(0·0438)	(0·0425)	(0·0461)
Proportional effect	0·1163	0·1077	0·1268	0·1030	0·1011
	(0·0324)	(0·0352)	(0·0408)	(0·0366)	(0·0395)
III. White only					
Absolute effect	0·1394	0·1296	0·1376	0·1417	0·1402
	(0·0216)	(0·0219)	(0·0221)	(0·0214)	(0·0221)
Proportional effect	0·1064	0·0949	0·1036	0·1049	0·1035
	(0·0214)	(0·0220)	(0·0220)	(0·0216)	(0·0221)
Other controls (all regressions in column):					
Lags	X	X	X	X	X
School variables[c]		X	X	X	X
Threats to learning			X	X	X
Student demographics				X	X
Census division dummies					X

[a]Figures in parentheses are standard errors. Sample size is 1809 for panels I, 1061 for panels II, and 5244 for panels III.

[b]All regressions include the student's 10^{th} -grade test score, the student's school's mean 10^{th} -grade test score, and a dummy indicating whether those variables are missing.

[c]Includes dummy variables for fraction of students black, fraction of students who drop out, fraction of class taking remedial math, and miles between high school and four year college/university when missing.

Table 14: Estimated effects of grading standards on high school graduation-semilog model[a,b]

Grading Standard	(1)	(2)	(3)	(4)	(5)
A. Intercept only model					
Absolute effect	−0·0013	−0·0006	0·0004	0·0040	0·0037
	(0·0052)	(0·0053)	(0·0017)	(0·0044)	(0·0017)
R^2	0·06	0·07	0·07	0·26	0·27
Incremental[c] χ^2		43·**	290·***	1575·***	105·***
		[20]	[18]	[82]	[9]
B. Growth and intercept model					
Absolute effect	0·0075	0·0058	0·0108	0·0082	0·0064
	(0·0108)	(0·0106)	(0·0106)	(0·0082)	(0·0080)
Proportional effect	0·0095	0·0072	0·0132	0·0077	0·0054
	(0·0111)	(0·0110)	(0·0110)	(0·0088)	(0·0083)
R^2	0·06	0·06	0·07	0·26	0·26
Incremental χ^2		45·**	63·***	1259·***	21·*
		[20]	[18]	[82]	[9]
Other controls (all regressions in column):					
Lags	X	X	X	X	X
School variables[d]		X	X	X	X
Threats to learning			X	X	X
Student demographics				X	X
Census division dummies					X

[a]Figures in parentheses are standard errors. Sample size is 8679.

[b]All regressions include the student's 10^{th} -grade test score, the student's school's mean 10^{th} -grade test score, and a dummy indicating whether those variables are missing.

[c]χ^2 statistics shown with degrees of freedom in brackets from a Wald test that the additional covariates are jointly no different than zero († p<0.10, * p<0.05, ** p<0.01, *** p<0.001).

[d]Includes dummy variables for fraction of students black, fraction of students who drop out, fraction of class taking remedial math, and miles between high school and four year college/university when missing.

Table 15: Estimated effects of grading standards on college attendance by 1984-semilog model[ab]

Grading Standard	(1)	(2)	(3)	(4)	(5)
A. Intercept only model					
Absolute effect	−0·0122	−0·0091	−0·0073	−0·0065	−0·0016
	(0·0085)	(0·0077)	(0·0077)	(0·0073)	(0·0072)
R^2	0·16	0·17	0·18	0·26	0·27
Incremental[c] χ^2		70·***	46·***	1158·***	77·***
		[20]	[18]	[82]	[9]
B. Slope and intercept model					
Absolute effect	−0·0183	−0·0179	−0·0131	−0·0109	−0·0046
	(0·0129)	(0·0126)	(0·0126)	(0·0120)	(0·0114)
Proportional effect	−0·0152	−0·0164	−0·0116	−0·0121	−0·0066
	(0·0122)	(0·0118)	(0·0119)	(0·0115)	(0·0107)
R^2	0·16	0·17	0·17	0·26	0·27
Incremental χ^2		62·***	46·***	1158·***	89·***
		[20]	[18]	[82]	[9]
Other controls (all regressions in column):					
Lags	X	X	X	X	X
School variables[d]		X	X	X	X
Threats to learning			X	X	X
Student demographics				X	X
Census division dummies					X

[a]Figures in parentheses are standard errors. Sample size is 8305.

[b]All regressions include the student's 10^{th} -grade test score, the student's school's mean 10^{th} -grade test score, and a dummy indicating whether those variables are missing.

[c]χ^2 statistics shown with degrees of freedom in brackets from a Wald test that the additional covariates are jointly no different than zero († p<0.10, * p<0.05, ** p<0.01, *** p<0.001).

[d]Includes dummy variables for fraction of students black, fraction of students who drop out, fraction of class taking remedial math, and miles between high school and four year college/university when missing.

Table 16: Estimated effects of grading standards on high school graduation by race-semilog model[ab]

Grading Standard	(1)	(2)	(3)	(4)	(5)
A. Intercept only model					
I. Hispanic only					
Absolute effect	0·0127	0·0140	0·0143	0·0082	0·0046
	(0·0115)	(0·0119)	(0·0074)	(0·0072)	(0·0022)
II. Black only					
Absolute effect	−0·0087	−0·0068	−0·0032	−0·0055	−0·0147
	(0·0142)	(0·0156)	(0·0075)	(0·0141)	(0·0190)
III. White only					
Absolute effect	−0·0002	−0·0011	−0·0001	0·0065	0·0068
	(0·0063)	(0·0067)	(0·0029)	(0·0021)	(0·0062)
B. Growth and intercept model					
I. Hispanic only					
Absolute effect	0·0161	0·0103	0·0171	0·0066	−0·0009
	(0·0217)	(0·0200)	(0·0199)	(0·0183)	(0·0182)
Proportional effect	0·0151	0·0112	0·0231	0·0047	−0·0037
	(0·0201)	(0·0173)	(0·0170)	(0·0166)	(0·0166)
II. Black only					
Absolute effect	−0·0059	−0·0032	0·0111	−0·0213	−0·0261
	(0·0254)	(0·0275)	(0·0300)	(0·0260)	(0·0265)
Proportional effect	0·0158	0·0241	0·0365	0·0218	0·0140
	(0·0203)	(0·0202)	(0·0226)	(0·0176)	(0·0189)
III. White only					
Absolute effect	0·0136	0·0109	0·0138	0·0146	0·0152
	(0·0147)	(0·0156)	(0·0154)	(0·0125)	(0·0120)
Proportional effect	0·0132	0·0104	0·0138	0·0107	0·0107
	(0·0161)	(0·0171)	(0·0169)	(0·0135)	(0·0129)
Other controls (all regressions in column):					
Lags	X	X	X	X	X
School variables[c]		X	X	X	X
Threats to learning			X	X	X
Student demographics				X	X
Census division dummies					X

[a]Figures in parentheses are standard errors. Sample size is 1836 for panels I, 1095 for panels II, and 5273 for panels III.

[b]All regressions include the student's 10^{th} -grade test score, the student's school's mean 10^{th} -grade test score, and a dummy indicating whether those variables are missing.

[c]Includes dummy variables for fraction of students black, fraction of students who drop out, fraction of class taking remedial math, and miles between high school and four year college/university when missing.

Table 17: Estimated effects of grading standards on college attendance by 1984 by race-semilog model[ab]

Grading Standard	(1)	(2)	(3)	(4)	(5)
A. Intercept only model					
I. Hispanic only					
Absolute effect	0·0080	0·0121	0·0157	0·0119	0·0172
	(0·0123)	(0·0115)	(0·0074)	(0·0122)	(0·0027)
II. Black only					
Absolute effect	−0·0337	−0·0250	−0·0196	−0·0269	−0·0111
	(0·0211)	(0·0210)	(0·0160)	(0·0215)	(0·0105)
III. White only					
Absolute effect	−0·0075	−0·0119	−0·0104	−0·0013	0·0031
	(0·0094)	(0·0097)	(0·0036)	(0·0031)	(0·0091)
B. Growth and intercept model					
I. Hispanic only					
Absolute effect	−0·0029	−0·0008	0·0069	0·0032	0·0058
	(0·0210)	(0·0203)	(0·0205)	(0·0214)	(0·0219)
Proportional effect	−0·0022	0·0007	0·0081	0·0034	0·0043
	(0·0194)	(0·0186)	(0·0188)	(0·0212)	(0·0214)
II. Black only					
Absolute effect	−0·0514	−0·0458	−0·0321	−0·0388	−0·0299
	(0·0281)	(0·0326)	(0·0311)	(0·0307)	(0·0293)
Proportional effect	−0·0348	−0·0362	−0·0252	−0·0436	−0·0291
	(0·0233)	(0·0256)	(0·0247)	(0·0253)	(0·0235)
III. White only					
Absolute effect	−0·0043	−0·0139	−0·0103	0·0074	0·0127
	(0·0163)	(0·0170)	(0·0173)	(0·0163)	(0·0161)
Proportional effect	−0·0034	−0·0126	−0·0088	0·0068	0·0100
	(0·0162)	(0·0172)	(0·0173)	(0·0163)	(0·0159)
Other controls (all regressions in column):					
Lags	X	X	X	X	X
School variables[c]		X	X	X	X
Threats to learning			X	X	X
Student demographics				X	X
Census division dummies					X

[a]Figures in parentheses are standard errors. Sample size is 1746 for panels I, 1024 for panels II, and 5086 for panels III.

[b]All regressions include the student's 10^{th} -grade test score, the student's school's mean 10^{th} -grade test score, and a dummy indicating whether those variables are missing.

[c]Includes dummy variables for fraction of students black, fraction of students who drop out, fraction of class taking remedial math, and miles between high school and four year college/university when missing.

E Models that drop observations rather than impute

The following two tables show the effects of grading standards on 12^{th} grade test score, high school graduation, and college attendance for the full sample, by gender, and by race under different methods of accounting for students with a missing 12^{th} grade test score or that transferred high schools. These results are described in Footnote 14 on page 12. Columns one through three present the coefficients for the absolute effect and columns four through six show the coefficients for the differential effect. The 'Base' refers to the model that is shown in the main body of this text where imputations were made. The other columns present results from models that exclude from the outcome regressions either students that had a missing 12^{th} grade test score or that transferred high schools as indicated by the column heading. Note that in all results shown in this paper, such students were excluded from the first stage regression that determined grading standards.

Table 18: Estimated effects of grading standards on 12$^{\text{th}}$ grade test score with models that drop observations with a missing 12$^{\text{th}}$ grade test score or that transfered high school[a,b]

Outcome and Sample	Absolute effect			Differential effect		
	Base[c]	Require 12$^{\text{th}}$ grade test score	Require stayed at high school	Base	Require 12$^{\text{th}}$ grade test score	Require stayed at high school
A. 12th grade test score						
All	0·1621 (0·0176) [8589]	0·1723 (0·0186) [7914]	0·1708 (0·0180) [8144]	0·1281 (0·0162) [8589]	0·1363 (0·0177) [7914]	0·1338 (0·0175) [8144]
By Gender						
Female	0·1429 (0·0222) [4277]	0·1661 (0·0245) [3966]	0·1674 (0·0239) [4069]	0·1152 (0·0205) [4277]	0·1402 (0·0229) [3966]	0·1424 (0·0229) [4069]
Male	0·1794 (0·0235) [4312]	0·1886 (0·0266) [3948]	0·1826 (0·0252) [4075]	0·1395 (0·0227) [4312]	0·1449 (0·0259) [3948]	0·1395 (0·0248) [4075]
By Race						
Hispanic	0·1546 (0·0216) [1809]	0·1827 (0·0368) [1694]	0·1867 (0·0361) [1714]	0·1050 (0·0192) [1809]	0·1311 (0·0317) [1694]	0·1345 (0·0319) [1714]
Black	0·2607 (0·0477) [1061]	0·2609 (0·0487) [997]	0·2796 (0·0490) [979]	0·2045 (0·0430) [1061]	0·2083 (0·0446) [997]	0·2222 (0·0460) [979]
White	0·1615 (0·0204) [5244]	0·1628 (0·0215) [4803]	0·1566 (0·0203) [5016]	0·1312 (0·0197) [5244]	0·1329 (0·0216) [4803]	0·1278 (0·0202) [5016]

[a]Figures in parentheses are standard errors. Sample sizes are in brackets.

[b]All regressions include lags, school variables, threats to learning, student demographic, and census division dummy controls. For details on these controls see Appendix B.

[c]Base refers to the model with the imputations described in Section II.D and used in all the tables shown in the main text of this paper.

Table 19: Estimated effects of grading standards on high school graduation and college attendance by 1984 with models that drop observations with a missing 12$^{\text{th}}$ grade test score or that transfered high school[a b]

	Absolute effect			Differential effect		
Outcome and Sample	Base[c]	Require 12$^{\text{th}}$ grade test score	Require stayed at high school	Base	Require 12$^{\text{th}}$ grade test score	Require stayed at high school
B. High school graduation						
All	0·0107	0·0063	0·0170	0·0126	0·0070	0·0165
	(0·0082)	(0·0081)	(0·0083)	(0·0091)	(0·0087)	(0·0094)
	[8679]	[7820]	[8161]	[8679]	[7820]	[8161]
By Race						
Hispanic	−0·0019	0·0255	0·0283	−0·0094	0·0227	0·0266
	(0·0062)	(0·0208)	(0·0195)	(0·0061)	(0·0190)	(0·0179)
	[1836]	[1669]	[1720]	[1836]	[1669]	[1720]
Black	−0·0257	−0·0271	0·0115	0·0100	0·0048	0·0307
	(0·0175)	(0·0280)	(0·0295)	(0·0161)	(0·0247)	(0·0253)
	[1095]	[994]	[988]	[1095]	[994]	[988]
White	0·0188	0·0020	0·0049	0·0223	0·0099	0·0141
	(0·0044)	(0·0121)	(0·0129)	(0·0050)	(0·0133)	(0·0149)
	[5273]	[4743]	[5023]	[5273]	[4743]	[5023]
C. College attendance						
All	−0·0075	−0·0085	−0·0078	−0·0029	−0·0025	−0·0024
	(0·0119)	(0·0117)	(0·0115)	(0·0115)	(0·0111)	(0·0110)
	[8305]	[7539]	[7813]	[8305]	[7539]	[7813]
By Race						
Hispanic	0·0103	0·0200	0·0084	0·0105	0·0213	0·0126
	(0·0082)	(0·0216)	(0·0212)	(0·0077)	(0·0192)	(0·0191)
	[1746]	[1594]	[1636]	[1746]	[1594]	[1636]
Black	−0·0117	−0·0165	−0·0071	−0·0229	−0·0132	−0·0088
	(0·0366)	(0·0340)	(0·0332)	(0·0343)	(0·0285)	(0·0285)
	[1024]	[939]	[928]	[1024]	[939]	[928]
White	0·0008	−0·0130	−0·0145	0·0109	−0·0009	−0·0030
	(0·0144)	(0·0166)	(0·0155)	(0·0143)	(0·0166)	(0·0154)
	[5086]	[4611]	[4844]	[5086]	[4611]	[4844]

[a]Figures in parentheses are standard errors. Sample sizes are in brackets.

[b]All regressions include lags, school variables, threats to learning, student demographic, and census division dummy controls. For details on these controls see Appendix B.

[c]Base refers to the model with the imputations described in Section II.D and used in all the tables shown in the main text of this paper.

REFERENCES

Joshua Angrist and Jorn-Steffen Pischke. *Mostly harmless econometrics: an empiricist's companion*. Princeton University Press, 2009.

Julian R. Betts. The impact of educational standards on the level and distribution of earnings. *The American Economic Review*, 88(1):266–275, 1998.

Julian R. Betts and Jeff Grogger. The impact of grading standards on student achievement, educational attainment, and entry-level earnings. *Economics of Education Review*, 22(4):343 – 352, 2003.

Hans Bonesrønning. Can effective teacher behavior be identified? *Economics of Education Review*, 23(3):237 – 247, 2004. ISSN 0272-7757.

Nicholas Scott Cardell and Mark Myron Hopkins. Education, income, and ability: A comment. *The Journal of Political Economy*, 85(1):211–215, 1977.

Judith Chevalier and Glenn Ellison. Career concerns of mutual fund managers*. *Quarterly Journal of Economics*, 114(2):389–432, 1999.

Robert M. Costrell. A simple model of educational standards. *The American Economic Review*, 84(4):956–971, 1994.

Ronald G. Ehrenberg and Michael L. Bognanno. Do tournaments have incentive effects? *The Journal of Political Economy*, 98(6):1307–1324, 1990.

David N Figlio and Maurice E Lucas. Do high grading standards affect student performance? *Journal of Public Economics*, 88(9-10):1815 – 1834, 2004.

Uri Gneezy and Aldo Rustichini. Gender and competition at a young age. *The American Economic Review*, 94(2):377–381, 2004.

Uri Gneezy, Muriel Niederle, and Aldo Rustichini. Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics*, 118(3): 1049–1074, 2003.

Roger Koenker and Kevin F. Hallock. Quantile regression. *The Journal of Economic Perspectives*, 15(4):143–156, 2001.

Edward P. Lazear. Performance pay and productivity. *The American Economic Review*, 90(5):1346–1361, 2000.

Dean R. Lillard and Philip P. DeCicca. Higher standards, more dropouts? Evidence within and across time. *Economics of Education Review*, 20(5):459 – 473, 2001.

Kevin M Murphy and Robert H Topel. Estimation and inference in two-step econometric models. *Journal of Business and Economic Statistics*, 20(1):88–97, 2002.

Derek Neal and Diane Whitmore Schanzenbach. Left behind by design: Proficiency counts and test-based accountability. *Review of Economics and Statistics*, 92(2): 263–283, 2010.

J. Neyman and Elizabeth L. Scott. Consistent estimates based on partially consistent observations. *Econometrica*, 16(1):1–32, 1948.

Muriel Niederle and Lise Vesterlund. Do women shy away from competition? Do men compete too much? *Quarterly Journal of Economics*, 122(3):1067–1101, August 2007.

Gerald S. Oettinger. The effect of nonlinear incentives on performance: Evidence from econ 101. *Review of Economics and Statistics*, 84(3):509–517, 2002.

Joseph Price. Gender differences in the response to competition. *Industrial and Labor Relations Review*, 61(3):320–333, April 2008.