

NBER WORKING PAPER SERIES

THE IMPACT OF GRADING STANDARDS
ON STUDENT ACHIEVEMENT, EDUCATIONAL
ATTAINMENT, AND ENTRY-LEVEL EARNINGS

Julian R. Betts
Jeff Grogger

Working Paper 7875
<http://www.nber.org/papers/w7875>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2000

This research was supported by a grant from the American Educational Research Association which receives funds for its "AERA Grants Program" from the National Science Foundation and the National Center for Education Statistics (U.S. Department of Education) under NSF Grant #RED-9452861. Opinions reflect those of the authors and do not necessarily reflect those of the granting agencies or the NBER. We thank Heather Rose and Rosemary Hyson for excellent research assistance.

© 2000 by Julian R. Betts and Jeff Grogger,. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

The Impact of Grading Standards on Student Achievement,
Educational Attainment, and Entry-Level Earnings
Julian R. Betts, and Jeff Grogger
NBER Working Paper No. 7875
September 2000
JEL No. I2

ABSTRACT

Despite recent theoretical work and proposals from educational reformers, there is little empirical work on the effects of higher grading standards. In this paper we use data from the High School and Beyond survey to estimate the effects of grading standards on student achievement, educational attainment, and entry level earnings. We consider not only how grading standards affect average outcomes but also how they affect the distribution of educational gains by skill level and race/ethnicity. We find that higher standards raise test scores throughout the distribution of achievement, but that the increase is greatest toward the top of the test score distribution. Higher standards have no positive effect on educational attainment, however, and indeed have negative effects on high school graduation among blacks and Hispanics. We suggest a relative performance hypothesis to explain how higher standards may reduce educational attainment even as they increase educational achievement.

Julian R. Betts
University of California , San Diego
La Jolla, CA 92093
and Public Policy Institute of California

Jeff Grogger
School of Public Policy and Social Research
Room 3250, Public Policy Building
UCLA
Los Angeles, CA 90095
and NBER
jgrogger@ucla.edu

I. Introduction

Over the years, economists have produced a voluminous body of empirical research that has attempted to explain student outputs in terms of school inputs. The results of this "educational production function" literature generally point to the conclusion that measured inputs have limited effects on student outcomes (Hanushek 1986; Betts 1996a). In light of these pessimistic findings, it is surprising how little empirical work has been devoted to understanding how other aspects of the educational environment affect student behavior. In particular, given economists' general interests in incentive schemes, it is surprising how little empirical work has focused on educational incentives.

Such incentives have been the focus of a small theoretical literature on educational standards.¹ A common finding in that literature is that higher standards may help some students at the expense of others. Whereas higher standards may lead more motivated students to increase effort, they may cause others to give up as the standard moves beyond their reach. Thus policies involving higher standards could potentially have adverse distributional consequences.

Distributional concerns arise not only in theoretical treatments of the topic, but also in policy discussions. Although elected officials, teacher's unions, and business leaders have embraced the notion that students must be held to higher standards if they are to learn more in school (Applebome 1996; Mitchell, 1996; American Federation of Teachers, 1999; Betts and Costrell, forthcoming), concerns also have been raised that standards may not benefit all students equally. Perhaps the greatest issue is that higher

¹ See for instance Kang (1985), Becker and Rosen (1990), Costrell (1994), and Betts (1998).

standards might work against students from groups with historically poor performance, such as racial minorities and students in inner-city schools.

Betts' (1995a) recent empirical work provides some evidence on these issues. Using data from the Longitudinal Study of American Youth (LSAY), he has shown that higher grading standards improve the achievement of the average American high school student. Because the LSAY samples relatively few minorities, however, Betts is limited in what he can say about distributional issues. He finds that the effects of grading standards are significantly greater among stronger students than among weaker students, as measured by their GPA, but he is unable to estimate separate effects by race.

In this paper, we expand on the existing empirical work in two important ways. First, we explicitly consider the distributional consequences of higher standards for student achievement. We do this in two ways. First, we use quantile regression methods to estimate the effects of higher standards at different points in the achievement distribution. Second, we stratify the sample by race/ethnicity to analyze directly whether higher standards have adverse consequences for minorities.

We also analyze the effect of schools' grading standards on a number of educational outcomes. We estimate the effects of higher standards not only on test scores but also on high school graduation, college attendance, and entry-level earnings. This provides estimates of the effects of grading standards not only on student achievement, but also on other measures of educational success that may have long-run consequences for the student's well-being. More generally, it provides a more comprehensive view of the effects of higher standards.

The only other empirical study that we know of that addresses similar questions is Lillard and DeCicca (forthcoming). Rather than analyzing the effects of grading standards, however, they focus on the effects of higher graduation standards, defined as the number of courses that students must take in order to graduate. They analyze high school dropout rates and attrition rates using both state- and individual-level data. They test whether dropout rates change by the same amount in all states, regardless of the extent to which each state changed the number of courses required for high school graduation. Their paper finds that raising graduation standards leads to moderate increases in the dropout rate. To put it more precisely, in those states that raised graduation requirements, dropout rates did not decline as much as in other states over the same period.

II. Methods

A. Data

We analyze data from the Sophomore Cohort of the High School and Beyond (HSB) survey. This widely studied survey has many features that make it useful for our analysis. First, it provides data from a national probability sample of roughly 15,000 students from 1000 schools. The HSB began tracking these students in 1980, when they were sophomores in high school, and followed them through 1992, when they were roughly 28 years old. In addition to data on educational achievement and attainment, the HSB provides data on students' post-schooling earnings for the period 1989 to 1991.² Moreover, unlike the LSAY, the HSB substantially oversamples blacks and Hispanics.

² For this reason, we use the HSB rather than the more recent NELS. At the most recent (1994) follow-up, the NELS students were only 20 years old, and many were full-time students. Thus the NELS is unsuitable for an analysis of post-schooling earnings.

We use these data to construct our dependent variables. These include 12th-grade test scores, dummy variables for high school graduation and college attendance, and entry-level earnings. The high school graduation dummy is equal to one if the student received a high school diploma, but not if she received a GED. The college attendance dummy is equal to one if the student attended a four-year college at any point during the first two years after leaving high school. Our earnings measure is the (logarithm of) average annual earnings, where the averaging is done over all years between 1989 and 1991 during which the student indicated that she worked at least nine months out of the year. We restricted the earnings variable in this way to focus on full-time, post-schooling earnings. We average over all years that satisfy our inclusion criteria in order to reduce the considerable sampling error that is present in entry-level earnings data. Our dependent variables are summarized in panels A-C of table 1.

The final virtue of the HSB is that it provides all the information we need to construct measures of grading standards for each high school. In 1980 and 1982, HSB respondents took nationally standardized tests in fields such as mathematics, science and English. In 1984, the administrators of the HSB conducted a high school transcript survey, obtaining the complete high school transcripts for nearly all students in the cohort. The level of detail provided in both the standardized tests and the transcript data permit us to construct detailed measures of each school's grading standards. We focus on schools' grading standards for math courses in light of findings by Grogger and Eide (1995) and Murnane, Willett and Levy (1995) that the impact of mathematics achievement on earnings both is substantial and has grown over time.

B. Measuring Schools' Grading Standards

Constructing grading standards requires two pieces of information: each student's standing relative to other students in his/her school, as measured through grades, and each student's standing relative to all students nationwide, as measured through test scores. A school's grading standard is simply a measure of how stringently it grades its students, relative to an objective measure of student achievement obtained from standardized tests. In the simplest terms, if one school generally gives B's to students who score in the 75th percentile on the nationwide test, whereas a second school generally gives C's to such students, then the latter school has higher grading standards. Formally, we estimate the link between test scores and grading standards by regressing math test scores on the student's grade point average (GPA) in math. Including separate intercepts for each school provides us with our estimated grading standards.

To see this, consider the following regression of the Grade 12 math test score A_{ij} for person i in school j :

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

where the α , β , and γ terms are parameters to be estimated, ε_{ij} is an error term, and n^{SCH} and n^{COR} equal the number of schools and the number of distinctly identified math courses within the data, respectively. The GPA variable indicates the student's average letter grade in math classes, the SCHOOL dummy variables represent separate intercepts for each school, and ρ_{ijm} is the *number* of math courses of a given type m taken by the student. The latter are meant to control for the fact that if somebody takes a more

difficult course, for instance algebra instead of business math, then we would expect him to obtain a higher test score, holding constant his GPA. Similarly, the total number of math courses taken should positively influence the test score, holding constant GPA.

If all schools graded in the same way, then the relation between test scores and letter grades should be identical across schools. In other words, all schools should have an identical intercept:

$$(2) \quad \alpha_1 = \alpha_2 = \dots = \alpha_{n_{SCH}}$$

Assuming that condition (2) fails, that is, that grading standards vary, a school with a higher α_j has high grading standards. To see this, consider the case of two schools. If $\alpha_1 > \alpha_2$, then students in school 1 receive higher standardized test scores, on average, than students in school 2 who earn the same grades. For example, this means that students with a B average at school 1 score higher than students with a B average at school 2. Thus, school 1 has higher grading standards. Ordinary least squares applied to equation (1) yields an estimated grading standard for each school in the sample, denoted by $\hat{\alpha}_j$, where the hat symbol denotes a statistical estimate.

In preliminary analyses we experimented with a number of different measures of grading standards. First, we added controls for the student's race, to guard against the possibility that the HSB test battery may be racially biased. Second, we added the square of GPA to control for possible non-linearities. Third, we added both the race controls and GPA squared. Fourth, we repeated each of these variants after substituting an omnibus measure of test scores for math test scores as the dependent variable in equation (1), while

at the same time replacing the student's math GPA with his overall GPA and likewise replacing the number of math courses with the total number of courses.

We found very high correlations between the models that controlled for race and the square of GPA and the models that did not. More to the point, the estimated effects of grading standards on student outcomes were generally similar regardless of how we estimated the grading standard. Likewise, the results were similar whether we used the math grading standard or the overall grading standard. In the analysis below, we focus on the math grading standards estimated without terms for race or GPA squared.

In all cases, the null hypothesis implied by equation (2) was rejected resoundingly, with p-values well below 5×10^{-6} . In other words, for students at different schools but with the same GPA and list of courses completed, achievement varied dramatically, indicating that grading standards vary significantly across American schools.³ Panel D of Table 1 summarizes the estimated grading standard and two other key explanatory variables.

C. Estimating the Effects of Grading Standards on Educational Outcomes

With the estimated grading standards at hand, we estimate the effects of grading standards on students' educational success by fitting regression models of the form:

$$(3) \quad y_{ij} = \delta \hat{\alpha}_j + X_{ij} \phi + Z_j \theta + u_{ij}$$

where y_{ij} is the outcome measure for the i th student in the j th school, X_{ij} is a vector of student background characteristics, Z_j is a vector of school characteristics, and u_{ij} is a disturbance term. The vector X_{ij} includes the students race/ethnicity, sex, family size, family structure, parental income, parental education, and in some specifications, the student's 10th-grade test score. The term Z_j includes the average 10th-grade test score of

³ Further details on the construction of the grading standards appear in Appendix 1.

students in the school; this variable appears in some specifications to control for peer effects and for possible reverse causation, as we discuss further below. The terms δ , ϕ , and θ are parameters to be estimated, where δ gives the effect of a unit change in the grading standard on the educational outcome.⁴

III. Results

A. Effects of grading standards on 12th-grade test scores

Table 2 reports regression estimates of the effects of grading standards on students' 12th-grade math test scores. One problem with 12th-grade test scores in the HSB is that only a fraction of the high school dropouts in the sample actually took the 12th-grade tests. To the extent that dropouts who took the test differ from their counterparts who did not, this may give rise to a sample selection problem.

The standard approach to this problem would be to estimate a two-equation sample selection model. The first equation would be used to predict whether or not the student took the exam. Predicted values from this selection equation would then be used to correct the regression of 12th-grade test scores on the grading standards.

The problem with the standard approach is that it presumes the availability of an instrumental variable that does not affect test scores but which can be used to predict who takes the test. It is difficult to conceive of such a variable, let alone find one among the variables included in the HSB. For this reason we take an alternative approach, which is similar to that adopted by Grogger and Neal (2000). The idea is to use the information at our disposal not to predict who is a test-taker, but rather to predict what the non-test takers would have scored if they had in fact taken the test. To do this we assume that the

⁴ With the exception of the quantile regressions, all standard errors below are calculated in a manner that accounts for the presence of multiple students per school.

students who did not take the 12th-grade test would have achieved the same relative score on their 12th-grade test as they achieved on their 10th-grade test. In other words, we compute the student's 10th grade percentile rank based on her 10th-grade test score, then impute to her a 12th-grade test score corresponding to that percentile rank in the distribution of 12th-grade test scores.

As discussed by Grogger and Neal (2000), this approach may imply some restrictive assumptions when the imputed scores are included in an ordinary least squares regression. When they are included in a quantile regression, however, the assumption under which the approach yields consistent estimates is that the imputation places the student on the correct side of the conditional quantile being estimated. This seems plausible, since we are in effect assuming that dropouts do not learn at a sufficient rate so as to advance their relative position in the distribution of test scores.⁵

We present three specifications that differ as to the background variables that are included in the model. In addition to the variables shown, all three specifications include an extensive set of controls for the student's race, sex, family structure, family size, parental occupation, family income, and residence in an urban, suburban, or rural location, as listed in the note to Table 2. Specification (2) adds to that list the student's 10th-grade math score. This provides an important control for the student's ex ante achievement.

Specification (3) further adds the school's mean 10th-grade math score. This provides a control for peer effects that may play an important role in both student learning and in the school's setting of its grading standard. Without controlling for mean test

⁵ Grogger and Neal (2000) take a somewhat different approach, imputing test scores to dropouts that are strictly below the conditional quantile in question. In practice, the approach we take here yields estimates that are similar both to those based on the Grogger-Neal approach and to those that ignore the sample-selection problem altogether.

scores, the models we estimate could suffer from reverse causation, which is an important form of endogeneity bias. A plausible hypothesis is that, all else equal, schools with higher-achieving students will set higher standards. If so, then the school's grading standard is correlated with average achievement in the school, and models that fail to control for average achievement suffer from endogeneity bias. Including average test scores controls for such bias and allows us to test whether higher-achieving schools indeed set higher standards.

Ideally, we would prefer to control for students' ex ante achievement using tests that were taken upon entry into high school. We would also prefer to control for the school average of student achievement at the start of grade 9. Such measures would be unaffected by the high school's grading standards. Unfortunately, the HSB allows us to control only for individual and school-average math achievement based on a test given near the end of grade 10. Thus, the grade-10 test scores may reflect the school's grading standard, leading us to underestimate the impact of grading standards on 12th-grade test scores. If so, then our estimates of the effects of grading standards that control for grade-10 test scores may be conservative compared to estimates that controlled for true ex ante achievement.

Turning to the estimates in table 2, comparing specification (1) to specification (2) in panel A reveals a pattern that appears in nearly all the regressions that we present. In the models estimated without controls for the student's ex ante achievement, grading standards appear to have positive and significant effects on educational outcomes. Including the student's 10th-grade test score greatly reduces this effect, however. In

panel A, the coefficient on the grading standard falls by more than half, from 0.510 to 0.225.

In specification (3), the coefficient on the school's mean 10th-grade test score is positive and significant, thus providing some evidence that is consistent with peer group effects. Even so, including the mean test score reduces the grading standard coefficient only slightly, which provides little support for the hypothesis that schools with higher-achieving students impose higher standards on average. In sum, all of the estimates in panel A suggest that higher grading standards are associated with higher 12th-grade test scores, on average.

How big are the predicted effects of higher grading standards? We can answer this in two ways. First, because one goal of imposing rigorous educational standards might be to reduce within-cohort variation in achievement, it makes sense to interpret the predicted effects in terms of the variation in achievement within a grade. Table 1 shows that the standard deviation of 12th-grade test scores is 10.08. The estimate in specification (1) of Table 2 indicates that a one-standard deviation increase in grading standards would raise twelfth-grade test scores by about 1.8 points, or almost two-tenths of a standard deviation, whereas the estimate in specification (2) implies an increase of only about 0.8 points, or less than one-tenth of a standard deviation. Thus, compared to the variance of test scores, the effects are quite small. Even a large increase in standards at low-performing schools would be unlikely to reduce the within-cohort variation in achievement by very much.

A second way to gauge the impact of raising standards is to compare the predicted gains with average gains in individual student test scores between grades 10 and 12. For

the subsample with valid test scores in both grades 10 and 12, the average achievement gain between 10th and 12th grades is about 2 points.⁶ The predicted gain in grade-12 test scores of 0.8 that arises from a one-standard deviation increase in grading standards amounts to a 40 percent gain in the mean rate of improvement between grades 10 and 12. In sum, the predicted effects of higher grading standards are small in terms of variation in achievement *among* students, but substantial in terms of the mean gain in student achievement.

Because of the importance of distributional considerations in discussions of educational standards, it is useful to go beyond averages and assess how grading standards affect achievement throughout the distribution of 12th-grade test scores. To do this, we report quantile regression estimates of the effects of grading standards for the 25th, 50th, and 75th percentiles of the 12th-grade test score distribution in panel B of table 2. Given the high significance of the students' 10th-grade test score in these regressions, we focus our discussion on the estimates from specifications (2) and (3).

Sub-panels 1 and 2 show that grading standards have similar effects at both the first quartile and the median of the 12th-grade test score distribution. In both cases, the estimate from specification (2) shows that a one-standard deviation increase in the grading standard leads to an increase of about 0.34 points on the 12th-grade test score. These estimates change little when we include the school's mean 10th-grade test score. Thus for the typical student, and the students below him in the achievement distribution, grading standards have a significant effect on achievement. Although these improvements would

⁶ As mentioned earlier, the mean test score in grade 10 that is shown in table 1 includes values set to zero for missing cases. Among students with non-missing test scores in both grade 10 and grade 12, the mean gain in test scores was about 2 points.

do little to equalize achievement among students, a 0.34 point gain does translate into a 17 percent gain in the average rate of learning between grades 10 and 12.

At the third quartile, grading standards have somewhat greater effects on student achievement. The estimate in specification (2) indicates that a one-standard deviation increase in grading standards increases 12th-grade test scores at the 75th percentile by 0.85 points. The estimate from specification (3) is essentially the same. Thus grading standards have different effects on students at different locations in the distribution of educational achievement. Higher grading standards have somewhat more than twice the effect on test scores at the top quartile of the distribution than they have at the median or bottom quartile. Thus higher standards may contribute to greater inequality in the distribution of educational achievement.

That said, it is important to keep in mind that we find increases throughout the distribution, rather than increases at the top and *decreases* at the bottom. Moreover, compared to the 10-point standard deviation in grade 12 test scores documented in Table 1, all of the estimated effects are rather modest in magnitude. Nevertheless, they do translate into meaningful increases in average individual rates of learning. For students at the 75th percentile, a one-standard deviation increase in standards is predicted to increase the gain in test scores by about 40 percent.

B. Effects Of Grading Standards On Educational Attainment

Although test scores provide one measure of educational success, educational attainment provides another that is arguably more important. Although recent research shows that students with higher test scores have higher wages, all else equal, than students with lower scores, the labor market payoff to finishing high school and attending college is

even greater. In table 3, we present estimates of the effects of higher grading standards on these two important measures of educational attainment.

The estimated effects of grading standards on educational attainment differ importantly from the estimated effects of grading standards on test scores. The estimate in specification (1) in panel A suggests that a one-standard deviation increase in grading standards would increase the high school graduation rate by 1.6 percentage points. Put differently, this amounts to an 8 percent decrease in the high school dropout rate. When the student's own test scores are included in the regression model, however, the estimate falls by 75 percent and becomes statistically insignificant. When we add the school's mean 10th-grade test score to the model, the grading standard coefficient becomes negative, although it remains small and insignificant. The college attendance results are similar.

Considering our results as a whole, we find higher grading standards to raise test scores, although we find them to have no significant effect on educational attainment. The estimates even suggest that higher standards could lead to slight decreases in high school graduation. Since one would generally expect higher test scores to lead to higher educational attainment, it is important to attempt to reconcile these apparently contradictory findings.

Most theoretical models show that failure rates may rise after a standard is raised, because the higher standard may induce some students to give up and exert less effort. Thus the notion that higher grading standards might lead to lower high-school graduation rates is not inherently at odds with theory. The question is how a rise in grading standards could produce a zero or even negative effect on graduation rates despite producing gains in test scores throughout the distribution of achievement.

The first point to note is that the effects of grading standards on test scores are largest at the 75th percentile. With a graduation rate of 80 percent, the 75th percentile student is already likely to be a probable graduate, so the test score effects at the 75th percentile are unlikely to produce an effect on graduation rates. At the other end of the distribution, one might imagine that large effects on test scores could lead to increases in graduation rates. At the low end of the distribution, however, the test score effects are smaller, which may again result in no effect on graduation rates.

Thus the finding of zero effects on graduation despite positive effects on test scores is fairly easy to explain. Because the estimates suggest a possible negative effect, however, albeit an insignificant one, it is useful to ask how a negative graduation effect could accompany positive test score effects throughout the distribution of test scores.

An explanation that is consistent with these results is based on the relative performance hypothesis. Because higher standards lead to higher gains for students near the top of the distribution than for students near the bottom, students near the bottom could perceive themselves as falling behind on a relative basis, despite their absolute gains. If students judge their likely success by comparing their performance with others rather than by evaluating their performance in absolute terms, then higher grading standards could lead students near the bottom of the distribution to quit school. Loury and Garman (1995) find evidence for a similar “relative performance” effect in their analysis of the impact of college selectivity and earnings. They find that college students whose own Scholastic Aptitude Test (SAT) scores are significantly below the median at their college are more likely to drop out of college than students with the same SAT score who are above the median at their college.

C. Estimated Effects Of Grading Standards On Earnings

Ideally, we would like to measure how higher grading standards in school affect the student's lifetime utility. An important component of utility is earnings, since earnings are an important determinant of consumption. In table 4 we present estimates of the effects of grading standards on (the logarithm of) earnings.

As desirable as it is to consider earnings effects, it must be said at the outset that the earnings data in the HSB suffer from a number of shortcomings. First, the earnings data pertain to the period 1989 to 1991, at which time the HSB respondents were typically 25 to 27 years old. Since entry-level earnings data are quite noisy, we average each student's earnings over all years in which she satisfied our sample inclusion criteria. Since our interest is in full-time earnings, we include only earnings data from years when the student reported working full-time for at least nine months out of the year. Finally, it is important to point out that our earnings samples are quite a bit smaller than the test score and educational attainment samples. Part of the sample size reduction stems from our sample inclusion criteria, but part of it stems from the fact that the HSB earnings data suffer from high levels of item non-response.

It is worth noting that the earnings equation does not control for the student's post-secondary schooling, occupational choices, or mobility. This affects the interpretation of our estimate of the effect of grading standards on students' post-schooling earnings. In principle, higher standards may affect a student's ultimate educational attainment, occupation, or mobility, and each of these factors may in turn affect earnings. In addition to these indirect effects on earnings, higher standards may exert an independent direct effect, affecting the student's earnings even controlling for

these other factors. By excluding these factors from the regression equation, we obtain “reduced form” estimates of the total effect of grading standards on earnings, that is, the sum of the direct and indirect effects.⁷

Results from the earnings regressions for the full sample are shown in panel A of Table 4. The estimate of the grading standards coefficient in specification (2) is marginally significant with a t-statistic of 1.63. It implies that a one-standard deviation increase in grading standards would increase earnings by nine-tenths of one percent. The estimate in specification (3), which includes the school's mean 10th-grade test score, is insignificant, but the grading standards coefficient itself does not vary that much between the two specifications.

One problem with analyzing the earnings of 25- to 27-year-olds is that the earnings function of those who attend college may differ considerably from that of those who do not. The reason is that the earnings of college attendees are observed just a few years after the worker finishes school, whereas the earnings of non-college attendees are observed after the worker has accumulated up to 7-9 years of work experience. To the extent that the higher achievement stemming from higher high school grading standards represents skill that is difficult for employers to observe initially, but which is revealed only over time, one might expect the effects of grading standards to be more apparent in the earnings of workers with 7-9 years of work experience than in the earnings of workers who have just entered the labor market.

To test this notion we split the sample by college attendance. The results for college attendees, shown in panel B, show little effect of grading standards on earnings, as expected. The results for non-college attendees, shown in panel C, are a bit more

⁷ Grogger (1996b) discusses this issue in greater detail.

favorable. The estimate in specification (2) has a t-statistic of 1.95 and indicates that a one-standard deviation increase in grading standards increases earnings by 1.4 percentage points. The estimate in specification (3) is slightly less significant, but slightly larger, and the coefficient on the school's mean 10th-grade test score is insignificant in any case. Thus the estimates suggest that higher grading standards may lead to small increases in earnings down the line, at least for students who do not attend college.

These results are consistent with the "learning by observation" story above. The result for non-college attendees is also consistent with predictions from Betts (1998). In that model, dropouts' earnings rise as standards increase because the marginal dropout is more skilled when standards are higher. As a result, higher standards raise the average skill of dropouts and hence their pay. For students who attend college, however, we cannot say for sure whether higher standards do not affect earnings, or whether the data at our disposal are simply inadequate for a decisive test. This question will have to be left for future work.

D. Race/ethnicity-specific estimates of the effects of grading standards

In table 5 we present race/ethnicity-specific estimates of the effects of grading standards on all of the various outcomes that we consider. In this table, each coefficient comes from a separate regression that contains all of the variables included in specification (3) in the tables above. Our motivation in providing these estimates is similar to our motivation in presenting the quantile regression estimates above. Distributional considerations are often at the fore in discussions of educational standards, and concerns about racial differences in the effects of higher standards are particularly acute in the eyes of many analysts.

Panel A presents estimates by race/ethnicity of the effects of grading standards on mean test scores. Grading standards have similar effects on whites and Hispanics. They also have positive and significant effects on blacks, but the estimated coefficient is smaller. That said, it is important to point out one important limitation of the race-specific estimates: because of the small sample sizes involved, many of the differences by race are not statistically significant, even though the coefficients themselves differ by a fair amount. This lack of statistical significance must be kept in mind in interpreting the inter-racial differences.

The quantile regression estimates appear in panel B. Once again, the estimates for whites and Hispanics are similar both to each other and to the pooled estimates in Table 2. They show that the effects of grading standards are greater at the upper end of the achievement distribution than at the lower end or in the middle. For blacks, however, the effects are nearly uniform across the distribution of test scores. They are smaller than the estimates for whites and Hispanics and are marginally significant as well. Grading standards appear to have less effect on the achievement of blacks than on that of whites and Hispanics.

High school graduation effects vary by race as well, though in a different way. Grading standards appear to have no effect on white graduation rates. The regression coefficient is minuscule and it is dominated by its standard error. For both blacks and Hispanics, however, grading standards have negative and significant effects on high school graduation. The results for Hispanics are most consistent with the relative performance hypothesis. The inter-quartile difference in the effect of grading standards on achievement is greater for Hispanics than for the other two ethnic groups. If students assess their

success on the basis of relative comparisons with peers of their own race and ethnicity, then we would expect higher grading standards to have the greatest negative effect on the graduation rate of Hispanic students.

At first glance, however, the relative performance hypothesis would appear not to explain the graduation effect for blacks. Of all racial/ethnic groups, the inter-quartile difference in the effect of grading standards on achievement is least for blacks. However, it seems plausible that blacks compare their performance not just to other blacks, but to the whites and Hispanics with whom they attend school. In this case, the relative performance hypothesis could potentially explain the negative graduation effect for blacks. Whatever the explanation, however, the contrast in graduation effects between whites and minorities lends credence to concerns that the winners and losers from higher grading standards might divide along racial lines.

In contrast to the graduation effects, the effects of grading standards on college attendance are insignificant for all three groups. The presence of insignificant college attendance effects despite significantly negative graduation effects suggests that those blacks and Hispanics who were dissuaded from graduating by higher grading standards would have been unlikely to attend college even if their grading standards had been lower.

The results from race/ethnicity-specific earnings regressions are presented in panels E through G. Of all the race/ethnicity-specific analyses, these are the most limited by small sample sizes. None of the estimates are statistically significant. The general patterns are similar to those in Table 4, however. The coefficients in panel F, pertaining to college attendees, are of mixed sign and highly insignificant. The coefficients in panel G,

pertaining to non-college attendees, are also insignificant. Like the pooled-sample results in Table 4, however, all of the estimates for non-college attendees are positive.

IV. Conclusions

Our results provide mixed evidence as to the effects of higher grading standards. The achievement results suggest that students respond favorably to the incentives provided by higher standards. Test scores rise in schools with higher standards, although they rise more for students near the top of the achievement distribution than for students near the bottom. These findings are similar to those of Betts (1995a).

When it comes to educational attainment, however, there is no evidence that higher standards raise either high school graduation or college attendance. For minorities, in fact, there is evidence that higher standards actually reduce graduation rates. There is some evidence that higher standards raise students' post-schooling earnings, although this effect is limited to non-college attendees. Moreover, the earnings analysis generally is limited by small sample sizes and the availability of only entry-level earnings data.

The notion that higher standards could result in both winners and losers is consistent with previous theoretical work. The potential puzzle in our results is that higher standards could lead to lower graduation rates despite their positive effect on test scores throughout the achievement distribution. We note that this finding is consistent with a relative performance hypothesis, whereby students judge their success not in absolute terms, but relative to their classmates. Because higher standards lead to greater increases in test scores at the top of the distribution than at the bottom, students near the bottom may perceive themselves as losing ground and give up on graduating as a result. Loury and Garman (1995) find evidence for such an effect at the college level.

Whatever the ultimate explanation for our findings, we note that they are consistent with theory in terms of their complexity. Theory predicts that different students may react differently to higher standards, and that as a result, policies based on higher standards may have both winners and losers. Put differently, higher standards are unlikely to yield Pareto improvements in educational outcomes. Our results suggest that the complexity of students' responses to higher standards needs to be better understood by policymakers before higher standards become the cornerstone of educational reform policies.

V. Appendix One: Calculation of Grading Standards

Overall math grade point average (math GPA) was calculated from grades in each math course taken, using 0.33, 0.5 and 1 as weights based on whether the individual course was taken on a trimester, semester, or, more commonly, a year-long basis. Letter grades were converted to grade points using the following rules:

Grade Points Assigned	Corresponding Letter Grades
4.33	A+
4.0	A
3.7	A-
3.3	B+
3	B
2.7	B-
2.3	C+
2.0	C
1.7	C-
1.3	D+
1.0	D
0.7	D-
0	F of "Fail"
Missing	Pass, I,W,R

Reasonable variations in the grade points attributed to each letter grade led to only minor changes to average math GPA. These variables were also used to keep the counts of the number of

credits taken in each course, which are regressors shown in equation (1). In cases where a student failed a course, but re-took and passed the course later, the student was assigned the letter grade and credit from the successful attempt, and the fail was not used to calculate either GPA or the sum of credits received, on the grounds that it is the *final* level of success in a given course that is likely to have the greater impact on a student's achievement by grade 12 or success later in life. In cases where a student failed a course and did not repeat it successfully, the failure was calculated in estimating GPA, and the course was not included in the list of credits taken. In a few cases, transcripts recorded a grade of "pass", in which case we did not include the course in the GPA calculations but added it to the tally of the number of courses taken by the student.

The grading standard models as in equation (1) were performed using the subsample of students who attended public schools, did not switch high schools, and who had non-missing GPA and test-score data. The math grading standard model included 798 dummies for schools, and 46 counts of math courses taken. The model was run with 8182 observations. We eliminated 26 observations for students whose schools would be represented by less than 3 students. (On average, there were 10.25 students per school.)

VI. References

American Federation of Teachers, *Making Standards Matter 1999* (Washington, D.C.: American Federation of Teachers), 1999.

Applebome, Peter. "Education Summit Calls for Tough Standards to be Set by States and Local School Districts." *New York Times*, March 27, 1996a, p. B7.

_____. "A Corporate Agenda." *New York Times*, March 28, 1996b, p. A11.

Becker, William, and Rosen, Sherwin. "The Learning Effect of Assessment and Evaluation in High School." Discussion Paper 90-7, Economics Research Center, NORC, 1990.

Betts, Julian R. "Do Grading Standards Affect the Incentive to Learn?" 1995a. Manuscript, University of California, San Diego, Department of Economics.

_____. "Does School Quality Matter? Evidence from the National Longitudinal Survey of Youth." *Review of Economics and Statistics*, May 1995b, 77(2), pp. 231-50.

_____. "Is There a Link between School Inputs and Earnings? Fresh Scrutiny of an Old Literature," in Gary Burtless, ed., *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*. Washington, D.C.: Brookings Institution, 1996a.

_____. "Do School Resources Matter Only for Older Workers?" *Review of Economics and Statistics*, 1996b, 78(4).

_____. "The Impact of Educational Standards on the Level and Distribution of Earnings." *American Economic Review*, 1998a, (88:1), pp. 266-275.

_____. "The Two-Legged Stool: The Neglected Role of Educational Standards in Improving America's Public Schools", 1998b, *Economic Policy Review*, (4:1), pp. 97-116

_____. "The Impact of School Resources on Women's Earnings and Educational Attainment: Findings from the National Longitudinal Survey of Young Women." forthcoming, *Journal of Labor Economics*.

Betts, Julian R. and Robert M. Costrell, "Incentives and Equity under Standards Based Reform", in Diane Ravitch (Ed.), *Brookings Papers on Education Policy 2001*, forthcoming, (Washington, D.C.: The Brookings Institution).

Costrell, Robert M. "A Simple Model of Educational Standards." *American Economic Review*, September 1994, 84(4), pp. 956-71.

Grogger, Jeff. "Does School Quality Explain the Recent Black/White Wage Trend?" *Journal of Labor Economics*, April 1996a, 14(2), pp. 231-53.

_____. "School Expenditures and Post-Schooling Earnings: Evidence from High School and Beyond," *Review of Economics and Statistics*, November 1996b 78(4).

_____. "School Violence, Educational Attainment, and Teacher Pay," mimeo, February 1996c.

Grogger, Jeff, and Eide, Eric. "Changes in College Skills and the Rise in the College Wage Premium." *Journal of Human Resources*, Spring 1995, 30(2), pp. 280-310.

Grogger, Jeffrey, and Derek Neal. "Further Evidence on the Effects of Catholic Secondary Schooling." In William G. Gale and Janet R. Pack, eds., *Brookings-Wharton Papers on Urban Affairs*, vol. 1. Washington: Brookings Institution, 2000.

Hanushek, Eric A. "The Economics of Schooling." *Journal of Economic Literature* 24, 1986, 1141-1177.

Hanushek, Eric A., and Lori Taylor. "Alternative Assessments of the Performance of Schools: Measurement of State Variations in Achievement." *Journal of Human Resources* 25 1990, pp. 179-201.

Kang, Suk. "A Formal Model of School Reward Systems," in John H. Bishop, ed., *Incentives, Learning, and Employability*. Columbus, Ohio: National Center for Research in Vocational Education, Ohio State University, 1985, pp. 27-38.

Lillard, Dean R. and Philip P. DeCicca, "Higher Standards, More Dropouts? Evidence Within and Across Time," *Economics of Education Review*, (forthcoming).

Loury, Linda Datcher and Garman, David. "College Selectivity and Earnings." *Journal of Labor Economics*, April 1995 13(2), pp. 289-308.

Mitchell, Alison. "President Urges State Standards for Education." *New York Times*, March 28, 1996, pp. A1, A11.

Moulton, Brent R. "Random Group Effects and the Precision of Regression Estimates." *Journal of Econometrics*, 1986, 32, pp. 385-97.

Murnane, Richard J; Willett, John B; Levy, Frank. "The Growing Importance of Cognitive Skills in Wage Determination." *Review of Economics and Statistics*, 1995, 77(2, May), 251-266.

Table 1
Summary Statistics for Various Outcomes and Key Explanatory Variables

A. 12th-grade test score (<i>n</i> =9,543)		C. Log earnings	
Mean	14.46 (10.08)	Full sample (<i>n</i> =6,694)	9.93 (0.39)
25th percentile	5.34	College attendees (<i>n</i> =2,403)	10.04 (0.39)
50th percentile	13.46	Non-college attendees (<i>n</i> =4,072)	9.86 (0.37)
75th percentile	22.28		
B. Educational attainment		D. Key explanatory variables	
High school graduation (<i>n</i> =10,124)	0.80	Math grading standard	2.78 (3.49)
College attendance (<i>n</i> =9,831)	0.32	Student's 10th-grade test score	11.36 (9.67)
		School's mean 10th- grade test score	13.26 (4.35)

Note: Figures in parentheses are standard deviations. Sample for the statistics in panel D is the same as that for the statistics in panel A. Statistics corresponding to other samples are similar.

Table 2
Estimated Effects of Grading Standards on Test Scores

<u>A. Ordinary least squares (mean) regression estimates</u>			
Variable	(1)	(2)	(3)
Math grading standard	0.506 (0.037)	0.225 (0.024)	0.197 (0.025)
Student's 10th-grade math score		0.891 (0.007)	0.883 (0.008)
School's mean 10th-grade math score			0.061 (0.022)
R-square	0.24	0.71	0.71
<u>B. Quantile regression estimates</u>			
<u>1. 25th percentile</u>			
Variable	(1)	(2)	(3)
Math grading standard	0.436 (0.029)	0.102 (0.018)	0.094 (0.018)
Student's 10th-grade math score		0.911 (0.008)	0.909 (0.007)
School's mean 10th-grade math score			0.020 (0.018)
R-square	0.12	0.49	0.49
<u>2. 50th percentile</u>			
Variable	(1)	(2)	(3)
Math grading standard	0.651 (0.045)	0.108 (0.014)	0.099 (0.016)
Student's 10th-grade math score		1.003 (0.006)	0.995 (0.006)
School's mean 10th-grade math score			0.029 (0.015)
R-square	0.16	0.54	0.54
<u>3. 75th percentile</u>			
Variable	(1)	(2)	(3)
Math grading standard	0.713 (0.043)	0.243 (0.022)	0.227 (0.024)
Student's 10th-grade math score		0.952 (0.009)	0.947 (0.009)
School's mean 10th-grade math score			0.037 (0.021)
R-square	0.14	0.51	0.54

Notes: Sample size is 9,543. Figures in parentheses are standard errors. In addition to the variables shown, all regressions include race dummies, a sex dummy, urban and rural dummies, parental occupation dummies, parental education dummies, family income dummies, a two-parent family dummy, and dummies for the number of siblings in the household. Specifications (2) and (3) include a dummy indicating whether the student's 10th-grade test score was missing.

Table 3
Linear Probability Estimates of the Effects of
Grading Standards on Educational Attainment

<u>A. High school graduation</u>			
Variable	(1)	(2)	(3)
Math grading standard	0.0047 (0.0014)	0.0011 (0.0014)	-0.0023 (0.0015)
Student's 10th-grade math score		0.0095 (0.0005)	0.0086 (0.0005)
School's mean 10th-grade math score			0.0072 (0.015)
R-square	0.085	0.13	0.14
<u>B. College attendance</u>			
Variable	(1)	(2)	(3)
Math grading standard	0.0068 (0.0018)	0.0008 (0.0017)	-0.0013 (0.0018)
Student's 10th-grade math score		0.019 (0.0006)	0.018 (0.0006)
School's mean 10th-grade math score			0.0044 (0.0015)
R-square	0.16	0.26	0.26

Notes: Sample size is 10,124 in panel A and 9,831 in panel B. Figures in parentheses are standard errors. In addition to the variables shown, all regressions include race dummies, a sex dummy, urban and rural dummies, parental occupation dummies, parental education dummies, family income dummies, a two-parent family dummy, and dummies for the number of siblings in the household. Specifications (2) and (3) include a dummy indicating whether the student's 10th-grade test score was missing.

Table 4
Ordinary Least Squares Estimates of the Effects of
Grading Standards on Log Earnings

<u>A. Full sample</u>			
Variable	(1)	(2)	(3)
Math grading standard	0.0053 (0.0016)	0.0026 (0.0016)	0.0019 (0.0018)
Student's 10th-grade math score		0.0078 (0.0006)	0.0076 (0.0006)
School's mean 10th-grade math score			0.0014 (0.0015)
R-square	0.12	0.14	0.14
<u>B. College attendees</u>			
Variable	(1)	(2)	(3)
Math grading standard	0.0026 (0.0025)	0.0006 (0.0017)	-0.0008 (0.0025)
Student's 10th-grade math score		0.0070 (0.0010)	0.0066 (0.0010)
School's mean 10th-grade math score			0.0034 (0.0022)
R-square	0.11	0.13	0.13
<u>C. Non-college attendees</u>			
Variable	(1)	(2)	(3)
Math grading standard	0.0055 (0.0020)	0.0039 (0.0020)	0.0043 (0.0023)
Student's 10th-grade math score		0.0050 (0.0008)	0.0051 (0.0009)
School's mean 10th-grade math score			-0.0007 (0.0019)
R-square	0.12	0.13	0.13

Notes: Sample size is 6,694 in panel A, 2,403 in panel B, and 4,072 in panel B. Figures in parentheses are standard errors. In addition to the variables shown, all regressions include race dummies, a sex dummy, urban and rural dummies, parental occupation dummies, parental education dummies, family income dummies, a two-parent family dummy, and dummies for the number of siblings in the household. Specifications (2) and (3) include a dummy indicating whether the student's 10th-grade test score was missing.

Table 5
Estimated Effects of Grading Standards on Various Outcomes, by Race/Ethnicity

<u>A. Test scores: ordinary least squares (mean) regression estimates</u>			
Variable	White	Black	Hispanic
Math grading standard	0.220 (0.030) [6,487]	0.108 (0.049) [1,342]	0.198 (0.057) [1,425]
<u>B. Test scores: quantile regression estimates</u>			
<u>1. 25th percentile</u>			
Variable	White	Black	Hispanic
Math grading standard	0.126 (0.025) [6,487]	0.056 (0.036) [1,342]	0.060 (0.036) [1,425]
<u>2. 50th percentile</u>			
Variable	White	Black	Hispanic
Math grading standard	0.111 (0.023) [6,487]	0.056 (0.031) [1,342]	0.100 (0.041) [1,425]
<u>3. 75th percentile</u>			
Variable	White	Black	Hispanic
Math grading standard	0.210 (0.030) [6,487]	0.064 (0.050) [1,342]	0.277 (0.088) [1,425]
<u>C. High school graduation</u>			
Variable	White	Black	Hispanic
Math grading standard	0.0007 (0.0016) [6,688]	-0.0091 (0.0037) [1,499]	-0.0074 (0.0037) [1,583]
<u>D. College attendance</u>			
Variable	White	Black	Hispanic
Math grading standard	-0.0014 (0.0021) [6,530]	-0.0060 (0.0041) [1,442]	0.0013 (0.0040) [1,518]
<u>E. Log earnings, full sample</u>			
Variable	White	Black	Hispanic
Math grading standard	0.0011 (0.0020) [4,733]	0.0043 (0.0053) [807]	0.0023 (0.0041) [966]
<u>F. Log earnings, college attendees</u>			
Variable	White	Black	Hispanic
Math grading standard	-0.0009 (0.0030) [1,811]	0.0025 (0.0089) [300]	-0.0044 (0.0075) [251]
<u>G. Log earnings, non-college attendees</u>			
Variable	White	Black	Hispanic
Math grading standard	0.0027 (0.0027) [2,791]	0.0090 (0.0064) [471]	0.0071 (0.0050) [678]

Notes: Figures in parentheses are standard errors. Figures in brackets are sample sizes. In addition to the variables shown, all regressions include the student's 10th-grade test score, the school mean 10th-grade test score, a dummy indicating whether the student's 10th-grade test score was missing, a sex dummy, urban and rural dummies, parental occupation dummies, parental education dummies, family income dummies, a two-parent family dummy, and dummies for the number of siblings in the household.