Real Costs of Nominal Grade Inflation?
New Evidence from Student Course Evaluations

Philip Babcock
Department of Economics
University of California, Santa Barbara

March, 2009

ABSTRACT

College GPAs in the United States rose substantially between the 1960's and the 2000's. Over the same period, study time declined by almost a half. This paper uses a 12-quarter panel of course evaluations from the University of California, San Diego to discern whether a link between grades and effort investment holds up in a micro setting. Results indicate that average study time would be about 50% lower in a class in which the average expected grade was an "A" than in the same course taught by the same instructor in which students expected a "C." Simultaneity suggests estimates are biased toward zero. Findings do not appear to be driven primarily by the individual student's expected grade, but by the average expected grade of others in the class. Class-specific characteristics that generate low expected grades appear to produce higher effort choices—evidence that nominal changes in grades may lead to real changes in effort investment.

I. Introduction

A substantial rise over time in GPAs at colleges in the United States has been documented, debated, lamented and theorized about, but empirical evidence on effects of grade inflation has been sparse and theoretical predictions have been ambiguous. Recent research documents a large drop over time in study times at U.S. colleges. This decline occurred over the same period as the rise in GPAs. The time trends suggest a causal link between grades and human capital investment. Confounding influences, though, include the changing composition of the college-going population and unobserved changes in the post-secondary institutions they attended. Given growing alarm among educators about low levels of student effort and engagement in college, evidenced by a spate of well-publicized books[1], the answer is worth knowing: Do nominal changes in grading patterns translate into real changes in human capital investment?

The macro findings motivate an investigation at the micro level. To investigate the mechanism at lower levels of aggregation, this paper uses evidence from a 12-quarter panel of student course evaluations. Holding fixed instructor and course, classes in which students expect higher grades are found to be classes in which students study significantly less. Results indicate that average study time would be about 50% lower in a class in which the average expected grade was an “A” than in the same course taught by the same instructor in which students expected a “C.” Simultaneity would suggest that this estimate of effort decline, if anything, is biased toward zero. Findings do not appear to be driven primarily by the individual student’s expected grade, but by the average expected grade of *others* in the class. Class-specific factors, then, that generate low

[1] Bok (2005), Hersch (2005), Nathan (2005).

expected grades appear to produce higher effort choices, evidence that nominal changes in grades may lead to real changes in effort investment.

The remainder of the paper is organized as follows: Section II reviews the literature and fleshes out the motivation, Section III describes the empirical strategy and presents main empirical results, Section IV discuses the results in a broader context, Section V concludes.

II. Motivation

A. Grade Inflation

Evidence of college grades increasing over time comes from a number of sources. Using data on self-reported GPAs, Kuh and Huh (2002) find that average grades increased for all types of post-secondary institutions between the 1980s and 1990's, with the largest increases observed at research universities. For a modest sample of colleges, information about grades has been made publicly available by the institution, obviating the need for self-reports. These data also show average GPAs rising in the 1960s, flattening somewhat in the 1970s, and rising from the 1980's to the 2000's, most steeply for highly selective colleges. [2] At Harvard, for example, about 49% of grades awarded in 2001 were A's, whereas in 1985 the fraction of A's was less than a third.[3] Adelman (1996), using institutional data from nationally representative datasets[4], finds average grades rising between the 1980's and the 1990s (though not between the 1970s and 1980s). The National Postsecondary Student Aid Study (NPSAS) shows GPAs for full

[2] Rojstaczer, The Washington Post, Jan. 23, 2003.
[3] Healy, Boston Globe, Nov. 21, 2001
[4] National Longitudinal Study of the High School Class of 1972, High School and Beyond Longitudinal Study of 1980 Sophomores, National Educational Longitudinal Study of 1988.

time students rising from 2.83 to 2.97 between 1993 and 2004, with clear upward changes in GPA observed for all types of colleges. Given earlier work and recent NPSAS data, it would appear that average GPAs rose between the 1960s and 2004, except for a flat spell during the 1970s, with the increase amounting to about .15 points per decade.[5]

B. Grade Compression

Do rising GPAs mean anything in real terms? The term "grade inflation" may be an imperfect analogy. Rising grades differ from rising nominal prices because inflated grades compress the grade distribution and may be less informative signals of ability. An existing theoretical literature explores grade compression (e.g., Ostrovsky and Schwartz, 2003, and Chan, Hao, and Suen, 2005). In general in this literature, grade compression leads to less informative signals that help low ability students while harming high-ability students. While theoretical work offers intriguing characterizations of a possible relationship between grade compression and employer information about prospective workers, there would seem to be little systematic evidence that grade compression in undergraduate institutions has created problems for employers. More A's have been awarded in individual classes, but cumulative GPAs would still seem to allow employers to draw very fine distinctions. In the NPSAS (2000), only 14.5% of students in 2000 had a GPA of 3.75 or higher.

[5] Measures and methodologies are not directly comparable, but multiple sources support this figure as a ballpark estimate. Rojstaczer (2003) estimates that average grades rose by about .15 grade points per decade for the schools in his sample. Kuh and Hu (1999) find a higher rate of increase (about .27 grade points) in self-reported GPAs from the mid-1980s to the mid-1990s. Adelman shows a lower rate of increase (about .08 grade points) between the early 1980s and the early 1990s. NPSAS gives an increase of .14 from 1993 to 2004.

Grade compression appears prevalent in the debate among education professionals over grade inflation, despite a dearth of systematic evidence showing its effects.[6] However, a second channel may be of first-order concern. Models of grade compression view grade inflation through the lens of strategic interactions between schools and employers. As such, they abstract from the student effort decision. The next subsection describes evidence and theory on the relationship between rising grades and student effort—the channel that will be the focus of this paper.

C. Effort

In 2004, Princeton University undertook a large-scale initiative to end grade inflation. Materials that explained the need for reform included the following. “Grading done without careful calibration and discrimination is, if nothing else, uninformative and therefore not useful; at worst, it actively discourages students from rising to the challenge to do their best work.” The materials also included quotes from students who seemed to feel that easy grading made them work less: "If I get the same grade for my very best work that I get for work that is not my very best, I'm less motivated to try to stretch as far as I can." "I got a D in math and I deserved it. [N]ow I try to try harder."

The evidence, of course, is anecdotal. Moreover, there is no clear reason the effect should go in the direction indicated by the quotes, rather than in the opposite direction. It is not hard to imagine other student responses: “What was required to get an A was too difficult, so I decided to take it easy and get a C instead.” Models in Betts (1996) and

[6] For example, a chancellor at the University of Massachusetts, Amherst describes the main practical problem with grade inflation as follows: “As grades increasingly reflect an idiosyncratic and locally defined performance levels, their value for outside consumers of university products declines. Who knows what an “A” in American History means?” John V. Lombardi, June 3, 2005, http://www.insidehighered.com.

Costrell (1994) clarify the intuition. In these models, education increases the student's value of marginal product. Firms do not observe productivity but do observe whether a student met the education standard. Thus, individual students care only about the signal associated with educational attainment, whereas a social planner takes into account levels of human capital acquired and its distribution. When a standard rises, relatively high ability students increase effort to meet it, while others (those who had been marginally willing to meet the lower standard) give up and reduce effort. Whether the change in average effort level rises or falls depends on the relative magnitudes of the opposing effects on different parts of the distribution.

We argue then that the effect of grade inflation on college effort choices is an empirical question, not clearly answered in the existing literature. Betts and Grogger (2000) offer evidence from the High School and Beyond survey that stricter grading standards in secondary school improve test scores of students. Figlio and Lucas (2004) find that elementary school students subjected to higher grading standards appear to benefit academically. However, these studies do not involve college students and do not reference student time use. One study at the college level, Greenwald and Gilmore (1997), reports a negative correlation between expected grade and student effort in 3 consecutive quarters of course evaluation data from the University of Washington. The authors examined correlation matrices on class aggregated data and applied factor analysis in cross section. The analysis here uses a larger sample over a longer time period, incorporates both student-level and aggregated data, and relies on changes over time in the panel of observations for identification.

There is some suggestive evidence at the macro level. Aggregate time spent studying by full-time college students declined from about 24 hours per week in 1961 to about 14 hours per week in 2004 (Babcock and Marks, 2008). Concurrent trends of rising grades and falling effort are suggestive of a causal link, but may not tell the whole story. The composition of the college student population has changed over time as have characteristics of post-secondary institutions. We report the above time trends to motivate an empirical investigation at lower levels of aggregation.

III. Data and Results

A. Data

Student course evaluation panel data from University of California, San Diego allow for a controlled analysis of the relationship between grades and effort. Students in all courses in all departments completed in-class course evaluations in about the seventh week of a ten-week academic quarter. Data available publicly on the UCSD website contain summary statistics of survey responses at the class level. Appendix A contains a sample of the survey questions and data for one course, taken from the website. These data were used to construct a sample of 7,889 classes covering 2003-2007. In addition, raw files of the individual student responses were obtained from the student organization that conducts the surveys. Student responses were anonymous and individuals cannot be tracked over time; however, individual observations were matched to aggregate class data for 6,753 classes (or 86% of the total sample of classes). Table 1 shows descriptive statistics of course evaluations for 6,753 classes from 338,414 students taught by 1568 instructors across all departments, offered in 12 quarters between Fall, 2003 and Spring, 2007.[7] UCSD, with 17,891 students admitted out of 40,518 applicants in 2005, is assumed to be representative of large, selective public universities.

Among the survey questions asked was "What grade do you expect in this course?" Classroom level summary statistics for this response include the number of A's, B's, C's, D's and F's expected in the class, which the table displays as a fraction of the number of respondents in the class.[8] From these variables we calculated an average expected grade for the class. The average of expected grades in the average class was 3.44. Surveys also ask "How many hours a week do you spend outside of class studying for this course?" Students responded by choosing one out of eleven possible bins. The table shows the fraction of respondents whose answers fell in each bin. A weighted average of bin midpoints was used to compute average study hours per week. This is the primary dependent variable in the regressions that follow. Average study time in the average class was 4.9 hours per week. The data also contain information on the number of students in the class, the number of surveys returned as a fraction of the enrollment, the fractions of the class that took the course as part of their major and as a general education requirement, and the fractions of the class that were seniors, juniors, sophomores, freshmen, extension students, and grad students. These are used as control variables in the regression analysis. All regressions also include year-quarter dummy variables to capture time trends.

[7] We exclude summer session classes because they are offered over five week rather than ten week time spans. Thus, weekly study times are not directly comparable. We also drop classes with less than ten survey respondents, as recommended in the psychometric literature on course evaluations (see Cashin, 1995.) Including classes with less than ten respondents does not significantly alter the results.

[8] This question was added to surveys in Fall 2003. Thus the time frame for our primary analysis is restricted to quarters no earlier than this.

Figure 1, a first pass at the data, plots average study hours against average expected grade. Panel A reveals that classes with higher average expected grade feature lower study times. The same negative correlation is visible at various levels of aggregation. Courses featuring higher expected grades feature lower study times (Panel B), instructors whose students expect higher grades elicit lower study times (Panle C), and departments characterized by higher expected grades feature lower study times (Panel D). The negative correlation persists within departments, as well. Panels E and F, show scatterplots of study time and average expected grade for the Math and Economics departments, respectively. As might be expected from different disciplines, the departments vary in average grades and study times. Despite the differences between departments, however, each of the scatterplots in Figure 1 shows a clear negative correlation between study time and expected grade. Other departments produce similar plots. The negative slope persists no matter how the data are sliced.

B. Empirical Strategy

Perhaps counter to intuition, classes in which students study more are classes in which expected grades are lower. This section describes the empirical strategy for parsing the channels that may be driving this correlation. It is reasonable to assume that study times and expected grades are simultaneously determined, i.e., that differences in study time alter one’s expected grade, even as differences in expected grade alter one’s study time choice. Grades, G, are assumed to be a function of study time T, class specific factors S, and student specific factors X:

$$G=f(T,S, X) \qquad 1)$$

It is assumed that $f_T>0$. Increased time spent on a course leads to a marginally higher grade (or expected grade.)[9] Given this channel alone, patterns in the data seem puzzling, because cov(G, T) is observed to be negative. Study times also depend on course and individual characteristics that co-vary with expected grade.

$$T=h(S, X) \qquad 2)$$

The interpretation emphasized in the following analysis is that class expected grade proxies for class-specific difficulty or grading standard, an important component of S. A negative correlation between expected grade and study time suggests that more demanding grading elicits higher effort choices. The causal channel described in 1)—a source of positive rather than negative correlation between the measures—would work against this effect. Thus, the remainder of the paper focuses on 2), with the understanding that estimated coefficients in regressions of study time on expected grade will be biased upward toward zero, due to simultaneity. [10]

Class expected grade may also proxy for course or instructor specific characteristics, distinct from grading standards. Self-selection by students of courses and instructors makes pooled regressions difficult to interpret. Specifically, one could imagine a dataset generated according to the following process:

$$T_{ijct} = \alpha + \beta_1 S_{jct} + \beta_2 X_{ijct} + \mu_{jc} + \varepsilon_{ijct}, \qquad (3)$$

where T_{ijct} is the study time for student, *i*, taught by instructor, *j*, in course, *c*, during term or quarter, *t,* and X_{ijct} are individual characteristics. S_{jct} , are class specific characteristics that very over time for a given instructor and course, e.g., the instructor's grading

[9] See Stinebrickner and Stinebrickner (2008). The authors find strong evidence of a (positive) causal link between study time and grades, using random assignment of roommates at Berea College.
[10] Data available for the empirical analysis include expected grade but not actual grade. However, expected grade is arguably the relevant empirical measure here, as a grade or grading standard not observed by the student should not influence student time use.

standard. The error term has two separate components, one that persists across time for a given instructor teaching a given course, μ_{jc}, and another that varies over time, ε_{ijct}. In many of the regressions in section III, we aggregate up to the class level:

$$\overline{T_{jct}} = \alpha + \beta_1 S_{jct} + \beta_2 \overline{X_{jct}} + \mu_{jc} + \overline{\varepsilon_{jct}} \qquad (4)$$

It is easy to imagine cases in which the μ_{jc} are correlated with the explanatory variables. For example, μ_{jc} might capture differences in the effort choices students make when facing different types of courses or instructors. Alternatively, μ_{jc} might capture the effect of high ability students selecting into classes with specific instructors or subjects. The fixed effects estimator overcomes this difficulty by estimating the μ_{jc} along with the other parameters and is mathematically equivalent to subtracting group means of the right and left hand side:

$$\overline{T_{jct}} - \overline{T_{jc}} = \beta_1 (S_{jct} - \overline{S_{jc}}) + \beta_2 (\overline{X_{jct}} - \overline{X_{jc}}) + (\overline{\varepsilon_{jct}} - \overline{\varepsilon_{jc}}) \qquad (5)$$

When group means have been subtracted, the term μ_{jc} disappears from the regression. The fixed-effects estimate is thus robust to unobserved instructor-course effects that persist over time. Similar strategies are used to control for instructor fixed effects and course fixed effects, in separate regressions. However, the course-instructor fixed effects specification would seem most appropriate, as it allows identification of β_1 from variations over time in course-specific practices of given instructors. Standard errors are clustered at the group level (e.g., errors are clustered at the course-instructor level when course-instructor fixed effects are used.)

Peer effects, selection within course and instructor, and other possible sources of endogeneity will be addressed in more detail in subsections III.C, D, and E.

B. Results

Aggregate regressions of class average study time on class expected grade provide a point of departure. Columns 1 and 2 of Table 2, show pooled regressions of weekly study hours on average expected grade (and constant), with and without controls. Column 2 indicates that a decrease of one grade point in the average expected grade of the class is associated with an increase in average study time of 2 hours per week. Signs on the other coefficients would seem consistent with intuition. Average study times fall as classes get larger, though at a diminishing rate. When a class contains a higher fraction of students taking the course as a requirement for their major or for a general education requirement, study times are higher. Drawing inferences about study-time from the pooled regressions is problematic, due to the biases discussed above. Columns 3, 4, and 5 report the results of course-instructor fixed effects regressions, respectively. In the preferred specification, Column 4, a decline of one grade point in average expected grade is associated with an increase of .94 hours in average time spent studying, and the estimate is highly significant. Coefficients on enrollment cease to be significant, and the sign of general education reverses.

Responses to a change in the average grade could mask important effects across the grade distribution. It could be, for example, that raising the standard for receiving an "A" increases effort whereas raising the standard for other specific letter grades decreases effort choices. While we cannot analyze this directly, we can explore how changes in the fraction of class expecting specific letter grades may influence study times. Results in Table 2, Column 5 derive from a regression using these fractions as the regressors of interest (with fraction of the class expecting C's as the excluded category.) When a higher fraction of students expect a D or F, study time rises; when a higher fraction expect B's or A's, study time falls. The coefficients in Column 6 imply responses similar in magnitude to those in the analogous average grade fixed effect regressions: When expected grade falls from an A to a C for all students in a class (two grade points), average study time appears to rise by 2.5 hours.

Theoretical models imply that when a standard rises, some students increase effort while some decrease their effort. One might expect that because some students will "give up" in classes with stricter standards and lower course grades, enrollment will be systematically lower in these classes. This could mean that estimated study times in difficult courses are inflated. (Some students reduce effort to zero, drop out, and are not counted in the average.) Courses may be dropped without consequence in the first three weeks of the quarter. This is well before student course evaluations are handed out and completed.[11]

Table 3 displays regression results when enrollment is the dependent variable.[12] Column 1 of the table is the pooled regression. We do not attribute a causal relationship to this regression, but note that higher grades appear to be awarded in smaller classes. When course is held fixed in Column 2, however, the pattern reverses: Higher grades are associated with significantly larger classes. It appears, then, that even though lower grades are associated with large-enrollment courses, when the same course is taught by a more lenient instructor, significantly more students enroll. The regression of interest is

[11] Enrollment figures come from late in the quarter, well after students have had an opportunity to drop without penalty.

[12] Percentage of returned surveys is excluded from the controls these regressions, as the dependent variable is built into the measure (the denominator). However, results are robust to its inclusion.

summarized in Column 3. Theory predicts that when a class is harder, students on the margin may reduce effort or perhaps drop the course. The instructor-course fixed effects regression in Column 3 shows that when a given instructor in a given course awards higher grades, enrollment is higher. One could interpret this to mean fewer students drop the course (prior to the course evaluation period) if grading is easier. But the coefficient is quite small (about 1 student out of a class of 100) and statistically indistinguishable from zero.[13] We do not, then, see strong evidence of students giving up in classes with lower expected grades.

C. Selection

Within-instructor variation in expected grade could arise from random variation in characteristics of students taking the class, rather than from exogenous variation in instructor standards and grading practices. If students with innate tendencies to study longer tend also to expect lower grades (perhaps because they study to compensate for lower endowed abilities), then classes with higher average study times will feature lower average expected grades. One way to account for selection within course and instructor is to compare student-level to aggregate data.

In the regressions in Table 4, observations are at the student level. Column 1, a summary of the pooled regression, indicates that students who expect higher grades do study less. But this result could be influenced by different types of students selecting into courses or instructors of differing difficulty. Column 2 summarizes a class fixed-effects

[13] One could argue that grades do not capture class difficulty as well as a measure of "effort per unit grade." As a robustness check, we constructed "effort per unit grade" (dividing average study hours by average grade) and used it as the dependent variable in regressions analogous to those in Table 3. The findings were similar. Holding fixed the instructor and the course, enrollments were marginally lower for classes featuring higher effort per unit grade. The coefficient was small and statistically insignificant.

regression of individual study times on individual expected grade. The coefficient on individual expected grade is negative and significant. This effect is not the result of grading standards or other differences between classes, but of differences between students (because within class, students face the same instructor and the same grading standard.) Within class, a rise in expected grade of one grade-point is associated with a decline in study time of only .15 hours, less than 10 minutes. Column 3 summarizes a course-instructor fixed effects model. Again students who expect a higher grade study slightly less than those who expect a lower grade. The model summarized in Column 4 includes as a regressor the average expected grade of *others* in the class. The coefficient on this regressor is much larger in magnitude than the coefficient on individual expected grade.

Loosely speaking, a student who expects an A rather than a B studies about 10 minutes less, whereas a class filled with students who expect A’s rather than B’s features an average study time more than *50 minutes lower*. Individual optimism, confidence or ability—as captured by individual expected grade—explains little of the decline in study times. The expected grade of others in the class would appear to be the dominant factor. This variable captures important aspects of the common environment, such as changes over time in course-specific practices and standards of a given instructor. It would appear that practices or environments causing students to expect lower grades generate higher study times.

D. Peer Effects

An alternative to the grading-standard explanation is that the coefficient on others' expected grade in Table 4 captures a peer effect—that when a student is surrounded by more confident or able peers, she studies less. Naïve peer regressions are plagued with correlated unobservables. The interpretation emphasized here has been that the "peer" coefficient captures characteristics of a common environment, rather than a true peer effect. But peer effects cannot be ruled out. Results here imply a social multiplier[14] larger than 5. Arguably, this may be implausibly large for a peer effect. For example, Glaeser *et al* (2002) , in a setting that features random assignment of students to dormitory rooms, find little evidence of a social multiplier on college GPAs (i.e., they find little difference between individual and aggregate-level coefficients in regressions of college GPA on student characteristics). Though the setting here differs substantially, the point is that there would seem to be little evidence in the literature that peer ability is many times more important than own ability in predicting college outcomes.

E. Simultaneously Observed Student Ratings

Table 5 suggests that random variations in class composition, within course and instructor, explain at best a small portion of the effect of average expected grade on study times. Variation in class composition, however, could be nonrandom within course and instructor. An alternative explanation for the main findings above is that changes in instructor practices and student decisions to enroll may be jointly determined. Both the pool of potential students and the instructor herself may have access to the instructor's rating from the previous time she taught the course. A drop in ratings, for example, may

[14] Glaeser, Sacerdote, and Scheinkman (2002) define the social multiplier as the ratio of the group level coefficient on a regressor to the individual level coefficient (or the amount that a coefficient rises in the movement from individual to aggregate-level regressions.)

cause the instructor to alter grading standards. Simultaneously, the ratings drop may attract systematically different students to the course (those less attentive to ratings--and perhaps less willing to study.) This could cause expected grades to rise and study times to fall even if there were no direct incentive effect of grading standards on effort choices.

One way to account for this possibility is to limit the sample to identical courses taught by the same instructor in *consecutive quarters* of an academic year. Student ratings of instructors are not available to instructors or students until late (about 5 weeks) into the subsequent quarter. The argument is that ratings *not yet jointly observed* cannot generate changes between quarters in instructor practices or student enrollment choices. The coefficient on expected grade in the first differences model is -1.17, statistically significant, and close to the corresponding estimate (-.94) in the fixed effects regression of Table 2, column 4.[15] Simultaneously observed course ratings do not appear to drive the results.

IV. Discussion

A. Implications

Theory on education standards leaves open the direction of the effect of a rise in standards on average effort. The data offer little evidence of a string "discouragement" effect of higher standards on marginal students. It could be that there are few students at the margin.[16] It could also be that elasticity of effort for non-marginal students is high

[15] Full results from the first differences model, which have not been reported in the tables, are available from the author upon request.
[16] More formally, because the standard is not binary, there are numerous "margins," one at each boundary that demarcates a letter grade from the next highest grade. The enrollment regressions do not shed light on effects at these margins.

enough to dominate the negative effect of discouraged marginal students on average effort.

Given some strong assumptions[17], a back-of-the-envelope estimate of the influence of observed grade inflation on study time is possible. The estimate in Table 2, if indicative of a causal link, implies that a 1-point increase in expected grade may reduce weekly study time by about .94 hours. If full time students take 4 classes each quarter, then uniformly increasing nominal (expected) grades for all classes by 1 point yields a decline of 3.76 hours. The observed increase in grades between the 1960s and 2004 was about .15 grade points per decade. Thus, the change in study times that would result is .56 hours per decade or a decline of about 2.3 hours a week between the 1960s and the 2000s. This would explain about a quarter of the observed 10 hour decline in weekly study time.

B Extensions and Related Work

The question of why grade inflation occurred has been left open. There would seem to be little evidence that recent college students are better prepared or higher in ability than students in previous cohorts.[18] This suggests grading standards may have declined. Though a conclusive explanation for a decline in grading standards is beyond the scope of this paper, the course evaluation data provide a possible clue. At UCSD, courses that feature higher grades receive higher student ratings, instructors who elicit

[17] General equilibrium effects, of course, have not been considered. When standards rise in a given class, students may study more or less in their other classes (Becker, 1982, Lichty *et al*, 1978). The effect on overall effort is not clear. It may be possible to address general equilibrium concerns in future research, as data on total student study times (linked to academic histories) become available.

[18] Average scores on college entrance exams appear, if anything, to have declined since the 1980s. Moreover, a comparison of college-going students in the NLSY79 to their counterparts in the NLSY97 shows that the later college cohort is drawn from students lower in the ability distribution (based on AFQT and ASVAB percentile rankings.)

higher expected grades receive higher ratings, departments with higher expected grades receive higher ratings, and a given instructor in a given course receives higher ratings when students' expect higher grades.[19] Figure 2 illustrates this last stylized fact for instructors and courses in the economics department.

These patterns are of course suggestive of a perverse incentive. An extensive literature in education finds that grading leniency explains a relatively small fraction of variance in evaluations.[20] In this sense, the effect of grading standards on instructor ratings seems "small." However, it is not clear that this measure of explained variance contains much information about instructor incentives. Even if students reward instructors more for becoming better communicators than for becoming easier graders, it may be far easier to change grading standards than to learn to be a better communicator. If so, instructors may face strong incentives to lower standards. A more systematic theoretical and empirical treatment of these issues is a subject for future research.

V. Conclusion

Average GPAs rose substantially between the 1960's and the 2000's for all categories of post-secondary education in the U.S. As higher average grades were awarded to college students, time spent studying declined dramatically, suggesting that nominal grades may influence time spent studying. But does the link between grades and effort investment hold up in a micro setting? A 12-quarter panel of course evaluations

[19] It could be that students receive higher grades and rate instructors more highly when they have learned more. However, Weinberg et al (2006) offer evidence that casts doubt on this explanation. They find that student evaluations are positively related to current grades but uncorrelated with learning once current grades are controlled. See also Johnson, 2003, for a summary of closely related evidence on the positive correlation between lenient grading and favorable instructor evaluations.

[20] See Cashin(1995) and Cohen(1981) for surveys of the early literature.

from the University of California, San Diego provides evidence. Keeping fixed instructor and course, classes in which students expect higher grades are found to be classes in which students study significantly less. Results indicate that average study time would be about 50 percent lower in a class in which the average expected grade was an “A” than in a class in which students expected a “C.” Simultaneity suggests estimates are biased toward zero. Findings do not appear to be driven primarily by the individual student’s expected grade, but by the average expected grade of others in the class. Class-specific factors that generate low expected grades appear to produce higher effort choices. Students appear to study more when the study time required to earn a given grade is (or is perceived to be) higher—evidence that nominal changes in grades may lead to real changes in effort investment.

References

Adelman, C. (1996), Principal Indicators of Student Academic Histories in Postsecondary Education, 1972-2000. Washington, DC: U.S. Department of Education, Institute of Education Sciences, 2004.

Babcock, P. and M. Marks (2008), "The Falling Time Cost of College: Evidence From Half a Century of Time Use Data," mimeo, University of California, Santa Barbara.

Becker, W. (1982), "The Educational Process and Student Achievement Given Uncertainty in Measurement," *American Economic Review*, 72(1), 229-236.

Betts, J. (1996), "The Impact of Educational Standards on the Level and Distribution of Earnings," *American Economic Review*, 88(1), 266-275.

Betts, J. and J., Grogger, J. (2000) "The Impact of Grading Standards on Student Achievement, Educational Attainment, and Entry-level Earnings," NBER working paper 7875.

Bok, Derek (2005), *Our Underachieving Colleges: A Candid Look at How Much Students Learn and Why They Should Be Learning More*, Princeton, NJ: Princeton university Press.

Cashin, W. E. (1995). "Student Ratings of Teaching: The Research Revisited," IDEA Paper No. 32. Manhattan, KS: Kansas State University, Center for Faculty Evaluation and Development.

Chan, W., H. Li, and W. Suen (2005), “A Signaling Theory of Grade Inflation,” mimeo, University of Toronto.

Cohen, P. A. (1981), “Student Ratings of Instruction and Student Achievement: A Meta-Analysis of Multisection Validity Studies.” *Review of Educational Research,* 51*,* 281–309.

Costrell, R. (1994), “A Simple Model of Educational Standards,” *The American Economic Review*, 84, 956-971.

Figlio D. and M. Lucas (2004), “Do high grading standards affect student performance?” *Journal of Public Economics*, 88, 1815-1834.

Greenwald, A.G. and G.M.Gillmore (1997), "Grading Leniency is a Removable Contaminant of Student Ratings," *American Psychologist* 11, 1209-17.

Healy, P. (2001), “Harvard Figures Show Most of Its Grades are A's or B's,” The Boston Globe, November 21, 2001.

Hersch, Richard and John Merrow (2005), *Declining By Degrees: Higher Education at Risk,* Palgrave Macmillan.

Johnson., V. (2003), “Grade Inflation: A Crisis in College Education”, New York : Springer.

Kuh, G. and S.Hu (1999), "Unraveling the Complexity of the Increase in College Grades from the Mid-1980s to the Mid-1990s," *Educational Evaluation and Policy Analysis*, 21, 1-24.

Lichty, R., D.Vose, and J. Peterson, 1978, "The Economic Effects of Grade Inflation on Instructor Evaluation: An Alternative Interpretation," *Journal of Economic Education,* 10, 3-11.

Malkiel, N. (2004), "Report to Faculty on Grading Proposals," memo, Princeton University, www.princeton.edu/%7eodoc/grading_proposals/.

McKenzie, R., 1975, "The Economic Effects of Grade Inflation on Instructor Evaluations: A Theoretical Approach," *The Journal of Economic Education*, 6, 99-105.

Nathan, Rebecca. (2005), *My Freshman Year: What a Professor Learned by Becoming a Student,* Ithaca, NY: Cornell University Press.

Ostrovsky M. and M. Schwartz (2003), "Equilibrium Information Disclosure: Grade Inflation and Unraveling," mimeo, Harvard University.

Rojstaczer, S. (2003), "Where All Grades Are Above Average," *The Washington Post,* Jan. 28, 2003.

Stinebrickner, R. and T. Stinebrickner, 2008. "The Causal Effect of Studying on Academic Performance," *Frontiers in Economic Analysis & Policy*, Berkeley Electronic Press, vol. 8(1), 1868-1868.

Weinberg, B., B. Fleisher and M. Hashimoto (2007), “Evaluating Methods for Evaluating Instruction: The Case of Higher Education,” NBER Working Paper #12844.

APPENDIX I

ECON ECON 136 (A) SU04

Enrollment: 33 Questionnaire's Returned: 20

	Freshman	Soph	Junior	Senior	Grad	Extension	#Resp
Your Class Level is	0	1	5	12	0	0	18
	0%	6%	28%	67%	0%	0%	

	Major	Minor	Gen. Ed.	Elective	Interest	#Resp
Your reason for taking this class is	13	1	0	4	2	20
	65%	5%	0%	20%	10%	

	A	B	C	D	F	P	NP	#Resp
What grade do you expect in this class?	11	7	0	0	0	1	0	19
	58%	37%	0%	0%	0%	5%	0%	

	N/A	Strong Dis	Disagree	Neither	Agree	StrongAgr	#Resp	Mean	StDev
1 Instructor displays a proficient command of the material	0	0	0	1	10	8	19	4.37	0.58
		0%	0%	5%	53%	42%			
2 Instructor is well-prepared for classes	0	0	0	0	6	13	19	4.68	0.46
		0%	0%	0%	32%	68%			
3 Instructor's speech is clear and audible	0	0	0	0	5	14	19	4.74	0.44
		0%	0%	0%	26%	74%			
4 Instructor's explains the course material well	0	0	0	1	8	9	18	4.44	0.60
		0%	0%	6%	44%	50%			
5 Lectures hold your attention	0	0	1	0	9	9	19	4.37	0.74
		0%	5%	0%	47%	47%			
6 Instructor's lecture style facilitates note-taking	0	1	1	0	2	15	19	4.53	1.09
		5%	5%	0%	11%	79%			
7 Instructor shows concern for students' learning	0	0	0	2	7	11	20	4.45	0.67
		0%	0%	10%	35%	55%			
8 Instructor promotes appropriate questions/discussion	0	0	0	2	7	10	19	4.42	0.67
		0%	0%	11%	37%	53%			
9 Instructor is accessible outside of class	2	0	0	1	4	12	17	4.65	0.59
		0%	0%	6%	24%	71%			
10 Instructor starts and finishes class on time	0	0	1	0	4	15	20	4.65	0.73
		0%	5%	0%	20%	75%			
11 The course material is intellectually stimulating	0	0	0	3	9	8	20	4.25	0.70
		0%	0%	15%	45%	40%			
12 Assignments promote learning	0	0	0	4	6	9	19	4.26	0.78
		0%	0%	21%	32%	47%			
13 Required reading is useful	0	0	0	3	6	10	19	4.37	0.74
		0%	0%	16%	32%	53%			
14 This course is difficult relative to others	0	1	4	10	4	0	19	2.89	0.79
		5%	21%	53%	21%	0%			
15 Exams are representative of the course material	0	0	0	3	5	12	20	4.45	0.74
		0%	0%	15%	25%	60%			
16 I learned a great deal from this course	0	0	0	1	10	7	18	4.33	0.58
		0%	0%	6%	56%	39%			
Instructor Question A (Question 21)	0	0	0	1	0	0	1	3	0
	0%	0%	0%	100%	0%	0%			
Instructor Question B (Question 22)	0	0	0	1	0	0	1	3	0
	0%	0%	0%	100%	0%	0%			
Instructor Question C (Question 23)	0	0	0	0	0	0	0		
Instructor Question D (Question 24)	0	0	0	0	0	0	0		
Instructor Question E (Question 25)	0	0	0	0	0	0	0		

	0-1	2-3	4-5	6-7	8-9	10-11	12-13	14-15	16-17	18-19	20+	#Resp	Avg
17 Study Hours per week	0	4	9	1	2	3	0	0	0	0	0	19	5.55
	0%	21%	47%	5%	11%	16%	0%	0%	0%	0%	0%		

	Rarely1	Some2	Most3	#Resp
18 How often do you attend this course? (1=Very Rarely, 2=Some of the Time, 3=Most of the Time)	0	0	20	20
	0%	0%	100%	

	No	Yes	#Resp
19 Do you recommend this course overall?	2	18	20
	10%	90%	
20 Do you recommend this professor overall?	0	19	19
	0%	100%	

Table 1
Descriptive Statistics

	(student)		(class)	
	Mean	St. Dev.	Mean	St. Dev.
Study hrs/wk	4.79	3.73	4.94	2.24
Grade	3.38	0.66	3.44	0.26
Enrollment(100s)	1.80	1.15	1.01	0.92
Fraction Returned	0.60	14.93	0.62	0.17
General Ed	0.23	0.31	0.17	0.26
Major	0.60	0.33	0.60	0.33
Senior	0.31	0.27	0.41	0.29
Junior	0.28	0.18	0.28	0.18
Sophomore	0.19	0.20	0.16	0.17
Freshman	0.20	0.30	0.13	0.24
Extension	0.00	0.01	0.01	0.02
Grad	0.01	0.04	0.02	0.06
A's	0.47	0.18	0.52	0.20
B's	0.44	0.14	0.41	0.16
C's	0.08	0.07	0.07	0.08
D's	0.002	0.01	0.002	0.01
F's	0.002	0.01	0.002	0.01
Study hrs: 0-1	0.12	0.13	0.12	0.16
Study hrs: 2-3	0.32	0.14	0.31	0.16
Study hrs: 4-5	0.27	0.10	0.26	0.11
Study hrs: 6-7	0.12	0.07	0.12	0.08
Study hrs: 8-9	0.06	0.05	0.06	0.06
Study hrs: 10-11	0.05	0.05	0.05	0.06
Study hrs: 12-13	0.02	0.03	0.02	0.04
Study hrs: 14-15	0.01	0.02	0.02	0.03
Study hrs: 16-17	0.01	0.01	0.01	0.02
Study hrs: 18-19	0.00	0.01	0.00	0.01
Study hrs: >=20	0.01	0.05	0.02	0.06
#Classes	6753			
#Courses	1859			
#Instructors	1568			
#Course-Instructor Pairs	3766			
#obs	338414			

Fall 2003 - Spring 2007, excluding summers.

Table 2
Dependent Var: Average Study Time

	1	2	3	4	5
Grade	-2.44***	-2***	-.945***	-.937***	-
	(.178)	(.173)	(.139)	(.139)	
A's	-	-	-	-	-2.48***
					(.428)
B's	-	-	-	-	-1.92***
					(.471)
D's and F's	-	-	-	-	1.78
					(2.18)
Enrollment/100	-	-.859***	-	-.152	-.145
		(.129)		(.136)	(.135)
Enroll/100-Squared	-	.146***	-	.0147	.0144
		(.0342)		(.0272)	(.0271)
Returned	-	-.0382	-	.0613	.0817
		(.276)		(.167)	(.168)
General Ed	-	1.31***	-	-.705***	-.719***
		(.252)		(.228)	(.228)
Major	-	3.36***	-	.699***	.667**
		(.211)		(.266)	(.265)
Senior	-	.344	-	-.0347	-.0624
		(.329)		(.28)	(.281)
Junior	-	-1.22***	-	-.00177	-.0198
		(.421)		(.241)	(.241)
Freshman	-	-.0145	-	-.0327	-.0239
		(.269)		(.189)	(.19)
Extension	-	10.9***	-	2.49	2.51
		(1.69)		(1.67)	(1.68)
Grad	-	2.51**	-	.189	.163
		(1.02)		(.744)	(.746)
Course-Instr F.E.	-	-	X	X	X
Groups	-	-	3766	3766	3766
Observations	6753	6753	6753	6753	6753

Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%
All regressions include year-quarter dummies as covariates.

Table 3
Dependent Var: Enrollment(100s)

	1	2	3
Grade	-.134***	.126***	.0148
	(.0414)	(.0317)	(.0392)
General Ed	.933***	-.194**	-.0511
	(.0698)	(.0939)	(.101)
Major	1.01***	.193***	.201**
	(.0489)	(.0722)	(.0782)
Senior	-1.23***	-.399***	-.211***
	(.0597)	(.0668)	(.0723)
Junior	-.46***	-.251***	-.00463
	(.0864)	(.0712)	(.0774)
Freshman	.0555	.167***	.249***
	(.0757)	(.0648)	(.0748)
Extension	-1.41**	-2.12***	-.902**
	(.587)	(.361)	(.377)
Grad	-1.6***	-.616***	-.398**
	(.182)	(.172)	(.185)
Course F.E.	-	X	-
Course-Instructor F.E.	-	-	X
Groups		1859	3766
Observations	6753	6753	6753

Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%
All regressions include year-quarter dummies as covariates.

Table 4

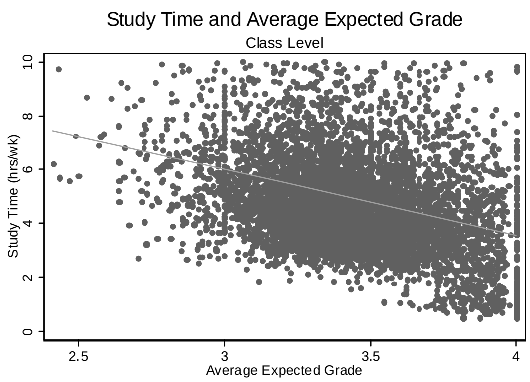
Dependent Var: Study Time (Individual)

	1	2	3	4
Grade (Indiv)	-.499*** (.0295)	-.146*** (.0118)	-.167*** (.0132)	-.166*** (.0133)
Grade (Class)	-	-	-	-.704*** (.0994)
Individ. Controls	-	X	X	X
Class F.E.	-	X	-	-
Course-Instr F.E.	-	-	X	X
Observations	338414	338414	338414	338414

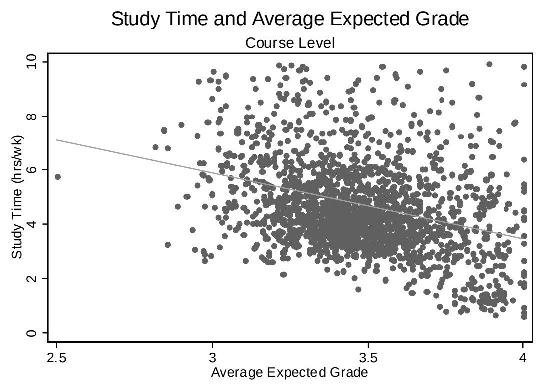
Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%
All models include time dummies, enrollment, and enrollment squared
as covariaites (except where these are collinear with fixed effects).

Figure 1

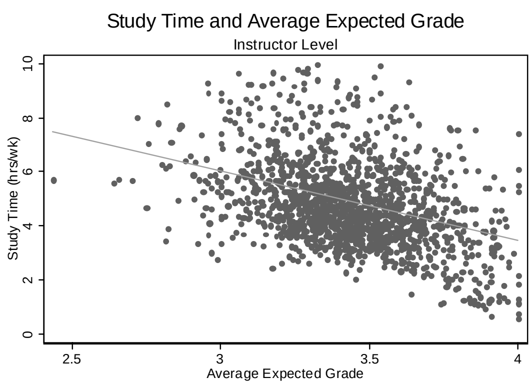
A.



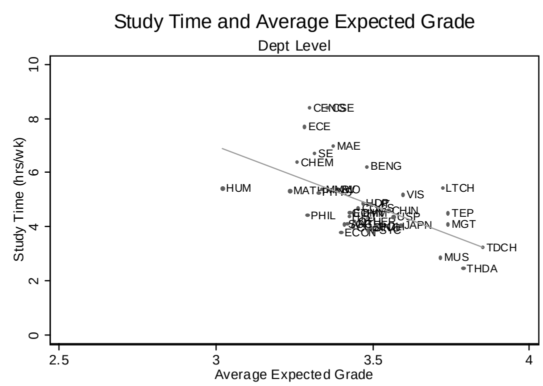
B.



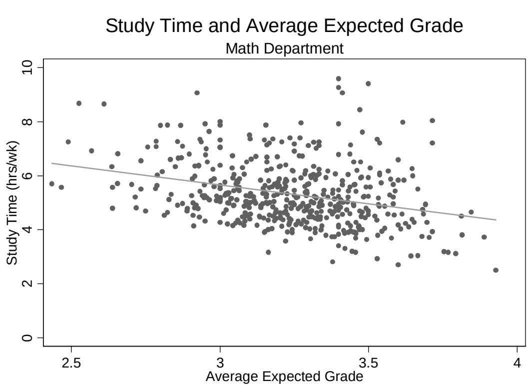
C.



D.



E.



F.

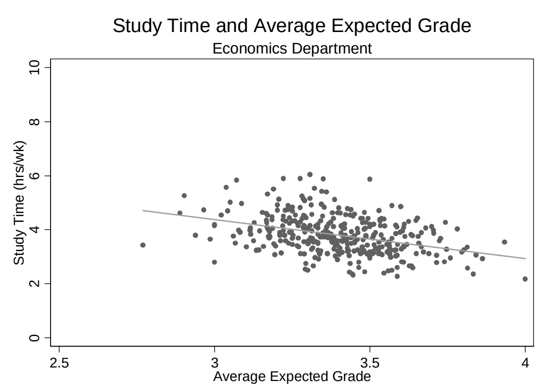
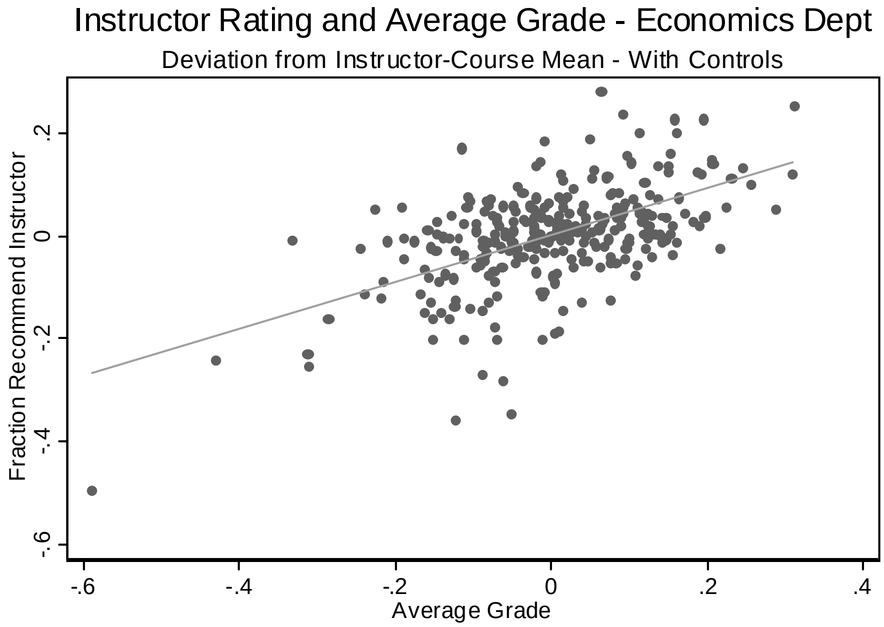


Figure 2


Instructor Rating and Average Grade - Economics Dept
Deviation from Instructor-Course Mean - With Controls
Fraction Recommend Instructor
.2
0
-.2
-.4
-.6
-.6
-.4
-.2
0
.2
.4
Average Grade