

The Gentleman's

New evidence that tough-grading teachers elicit better student performance

by DAVID N. FIGLIO & MAURICE E. LUCAS

WITH REPORTS THAT SOME OF THE NATION'S FINEST UNIVERSITIES have been handing out As like lollipops at Halloween, the lowering of standards in higher education has become a hot topic. But grading standards in primary and secondary education have received remarkably less attention.

There are two major questions related to grading standards. First, to what degree do the grades distributed by schools and teachers correspond to their students' performance on state and national exams? Second, and more important, how does "tough" or "easy" grading affect students' learning?

The literature on these questions is extremely thin. In fact, to our knowledge, the analysis presented here represents the first study to examine the grading standards of individual *teachers* and how those standards affect students' performance on independent exams. Our data set enabled us to examine the test-score gains of individual students from grade to grade across three school years. Thus we can see how individual students perform on nationally normed exams as they move from "tough" to "easy" grading teachers and vice versa. Our results suggest that elementary-school students learn more with "tough" teachers, with the effects varying depending on students' initial performance levels and on the overall performance level of their classrooms.

Measuring Grading Standards

For this study, we analyzed confidential data provided by the school board of

ILLUSTRATION BY IMAGES.COM



Alachua County, Florida, which includes the city of Gainesville. The data consist of observations on almost every 3rd, 4th, and 5th grader in the school system between the 1995–96 and 1998–99 school years, allowing us to follow two cohorts with three years of data each. Alachua County Public Schools is a relatively large district (by national standards), averaging about 1,800 test-taking students per grade, per year. The county is racially diverse, with a student population that is 60 percent white, 34 percent African-American, 3 percent Hispanic, and 2 percent Asian. Nearly half of all students are eligible for subsidized lunches, while 19 percent are identified as gifted, 8 percent as learning disabled, and less than 1 percent as English learners.

Alachua County provides a unique advantage for a study of this nature because it administers both the Iowa Test of Basic Skills (ITBS), a nationally normed exam, and the Florida Comprehensive Assessment Test (FCAT). The FCAT was designed to measure the degree to which students are meeting the Sunshine State Standards, which are also supposed to be the basis for students' letter grades. Today, the state uses the FCAT both to rate public schools as part of the "A+" accountability plan and to determine whether students in certain grades are promoted to the next grade. During the period covered by our data, however, these tests were used for informative and diagnostic purposes only (though they were publicly reported each year at the school level).

Students receive scores on the FCAT ranging from 1

(lowest) to 5 (highest), with the thresholds for each performance level designed to correspond with the letter grades A through F. Thus results from the FCAT are ideal for developing a measure of how generous individual teachers' grading policies are. We can then examine the relationship between teachers' grading standards and their students' performance gains on a different test, the ITBS.

Our primary measure of teachers' grading standards is the average gap between the letter grades given by particular teachers and the FCAT scores attained by their students. Students take the FCAT math exam in 5th grade and the FCAT reading exam in 4th grade. Consequently, this measure of grading standards is calculated using the math grades and test scores of 5th-grade teachers, and the reading grades and test scores of 4th-grade teachers. Examining students' performance on the ITBS in the 3rd, 4th, and 5th grades enables us to see how their gains in reading from 3rd to 4th grade, and in math from 4th to 5th grade, were affected by their teachers' grading standards that academic year.

Grade Inflation

It turns out that the grades teachers assign are highly correlated with students' ITBS and FCAT scores. But teachers also tend to grade far less stringently than the state standards indicate they should (see Figure 1). For instance, just 9 percent of students who were awarded A's by their teachers

Just 9 percent of students who were awarded A's by their teachers attained a score of 5 on the Florida Comprehensive Assessment Test.

attained a score of 5 on the FCAT. In fact, just 50 percent attained even a 4. Only 11 percent of students awarded B's by their teachers attained level 4 or above, and a mere 39 percent attained level 3 or above. And of students awarded C's, only 14 percent attained level 3 or above, and only 39 percent attained level 2 or above. Put differently, 86 percent of "C students" failed to achieve the minimum level of competency accepted (Level 3) on the Florida standards, along with 61 percent of "B students" and 17 percent of "A students." Yet not all teachers are so lax: Among the top half of teachers as ranked by their grading standards, 65 percent of A students attained level 4 or above while just 5 percent attained level 2 or below.

In short, teachers vary considerably in their grading standards, even within a single school district. In fact, teachers' grading standards often vary as much within a single school as within the school district as a whole. For instance, during the 1997–98 school year, the district-wide standard deviation in teacher-level grading standards was 0.68, while the mean within-school standard deviation in grading standards was 0.60. Tough-grading teachers, it appears, often teach alongside teachers with far lower standards. For research purposes, this is reassuring, since our empirical strategy relies mainly on within-school variation in teachers' grading standards to isolate the effects of those standards.

A Stable Trait

Estimating the effect of individual teachers' grading standards on their students' achievement gains assumes that these standards remain relatively consistent over time, that they are not unduly influenced by the composition of their class, and that they are not a reflection of some other observable characteristic that might account for any effects we observe. Fortunately, our data provide evidence in support of each of these assumptions.

To see whether teachers' grading standards remained stable over time, we divided the full sample of teachers into thirds according to their grading standards each year and examined how the position of individual teachers changed from year to year. For instance, we found that 75 percent of the teachers whose standards put them in the "easy" category (on a scale from "easy" to "moderate" to "tough") in one year remained in that category the following year, while just 6 percent evolved from easy to tough graders in a single year. This trend was essentially the same across the three categories,

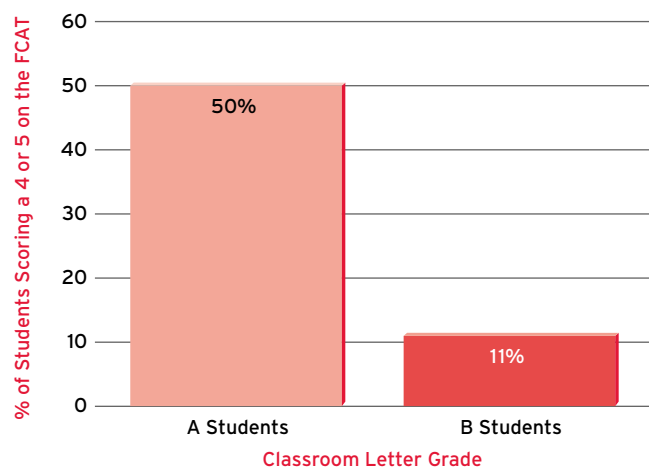
with very little movement between groups.

Nor does it appear that teachers' grading standards are influenced by the ability level of their students. To gauge this, we compared teachers who taught a higher ability class, as measured by their average 3rd-grade test scores, in 1998–99 than in the previous year, and vice versa. We found that even large changes in the ability level faced by teachers do not seem to affect their grading standards.

Turning finally to the relationship between other observable teacher characteristics and grading standards, we found that relatively tough graders are in fact slightly more experienced and slightly less likely to have attended a selective or highly selective undergraduate institution, though these differences are not statistically significant. However, tough graders are significantly more likely to hold Master's degrees. In any case, our analysis below controls for each of these measures of teachers' qualifications in order to rule out the possibility that teachers' observed characteristics drive the estimated effects of grading standards on student outcomes.

Grade Inflation (Figure 1)

Overall, students who receive high grades tend to attain higher scores than their peers on the Florida Comprehensive Assessment Test (FCAT). However, just 50 percent of "A students" and 11 percent of "B students" received scores of 4 or higher on the FCAT.



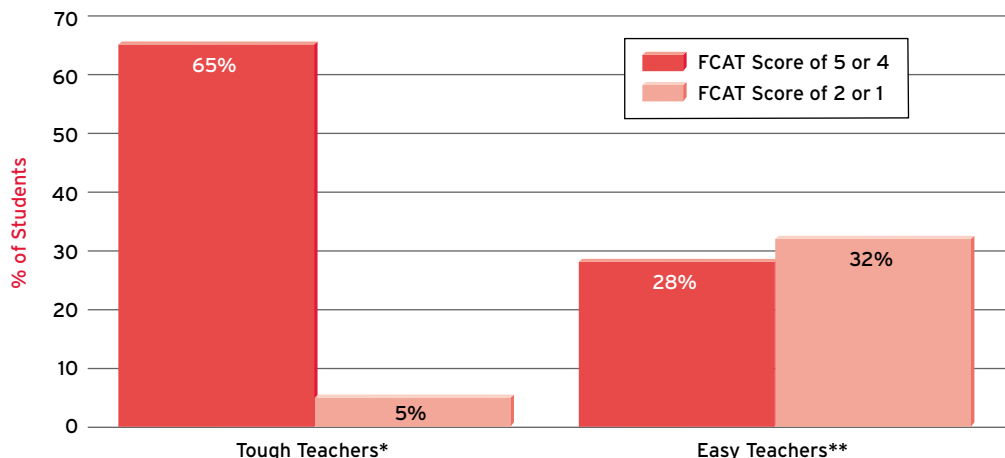
Note: FCAT scores are designed to correspond to letter grades, e.g., 5 = A, 4 = B.

SOURCE: Authors

“Tough” versus “Easy” Teachers (Figure 2)

Students who earn an A from a teacher with high grading standards are more than twice as likely to score a 4 or better on the FCAT as are students who receive A’s from an easy-grading teacher.

Percentage of “A Students” Scoring a 4 or Above or a 2 or Below on the FCAT in Classrooms with Different Grading Standards



* Teachers whose grading standards are above the median.

** Teachers whose grading standards are below the median.

Note: FCAT scores are designed to correspond to letter grades, e.g., 5 = A, 4 = B.

SOURCE: Authors

Classroom Assignment

The method by which students are assigned to teachers could also cause problems for an analysis of the effect of grading standards on performance gains. The fact that students are not assigned to teachers randomly would be especially troublesome in a cross-sectional analysis, in which one compares one classroom with another in the same year. For instance, looking in cross-section across our own data set reveals that teachers with high standards also have students who are more likely to be white or gifted and less likely to be low-income or learning disabled. This is true even within a given school. Hence it is unclear whether the outcomes associated with high standards are actually due to the standards themselves.

But our analysis looks at year-to-year changes in the grading standards faced by a student, making this less of a concern. We found that students are nearly as likely to move to a teacher with different standards as to experience the same grading standards from year to year. For instance, just 57 percent of students with teachers whose grading standards are below the median within their own school continue to have below-median teachers the next year. Likewise, 54 percent of students with above-median teachers continue to have above-median teachers the next year. This indicates that

year-to-year differences in grading standards within schools are close to random. Similar patterns are observed for most sub-groups—black and white students are approximately equally likely to move between groups, as are students who are eligible and ineligible for a free lunch. It is the case that gifted students (around 12 percent of the sample), no matter where they start out, are considerably more likely to be placed with a high-standards teacher the next year than are nongifted students. Nevertheless, the vast majority of students are almost as likely to move between low-standards and high-standards teachers as to experience the same level of standards across years. And the results presented below do not change materially when

gifted students are excluded from the analysis.

Empirical Results

For our primary analysis, we controlled for the average annual gain made by all students in the relevant school during the period of analysis; for such classroom characteristics as the share of white students, the share eligible for free lunches, and the students’ average math score in 3rd grade; and for the teacher’s years of experience, education level, and the selectivity of his or her undergraduate institution. In the end, we were interested in the effects on ITBS scores of changing a student from one level of grading standards to another.

Using this strategy, we found statistically significant improvements in test scores associated with higher standards. An increase in measured grading standards of 1 standard deviation is associated with about one-fifth of a year of schooling’s worth of gains in test scores—a large effect relative to many other interventions. (A year of test-score gain is measured as the average gain from one year to the next in Alachua County Public Schools. Because Alachua County’s gain scores tend to be larger than the national average, these are more conservative estimates of years of gain than are those

based on national grade equivalents.)

While the average effects of grading standards are important, the theoretical literature on grading standards suggests that higher standards could produce both winners and losers. Students who achieve a given standard may be made better off because the standard becomes a more meaningful accomplishment. But those students who are not able to achieve the standard precisely because it is now more rigorous are made worse off.

To study this issue, we tested whether the effect of high grading standards varied for students with different initial test scores in 3rd grade. We found that the positive effects of higher grading standards were restricted to those students who were no more than 0.8 and 0.9 standard deviation below the average score in reading and math, respectively. For students performing below this threshold, grading standards had no detectable effect, either positive or negative, on performance. We also found that higher-achieving classes, as measured by their average 3rd-grade test score in the relevant subject, may fare somewhat better than lower-achieving classes under teachers with tough grading standards.

Equally interesting, however, are the distributional effects of tough grading standards on performance gains within a given classroom. Put differently, are the benefits of high standards uniform within a class, or do some children benefit more than others? We found that high-achieving students benefit most from tough grading standards when they are placed in classrooms where the overall level of achievement is relatively low (see Figure 3). The opposite is also true: tough grading standards elicit the most improvement from low-achieving students when they are in classrooms with relatively high overall achievement.

This result has intuitive appeal. Since the grades assigned vary much less across classrooms than does students' performance on standardized tests, high-achieving students should be more likely to earn high grades in classrooms where the other students, on average, do not perform well on external assessments. Likewise, low-achieving students in classes with many strong students run a greater risk of receiving a low grade than they do when in low-achieving classes. So it seems only sensible that initially high-achieving students benefit most from tough teachers when they are among the strongest members of a class. Similarly, initially low-achieving students are challenged more to get a good grade with tough teachers, but particularly when they are among the weakest members of a class.

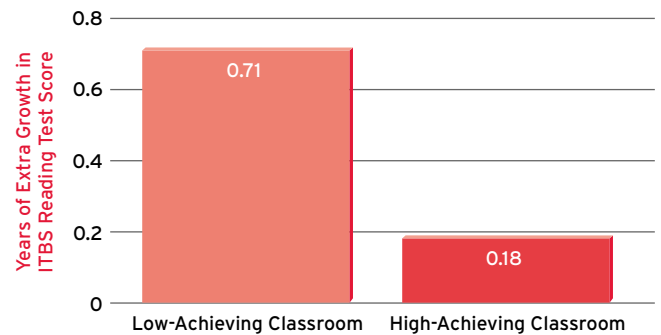
Parental Involvement

What might explain the positive effects of higher grading standards? One possibility, of course, is that high standards motivate students to work harder. A second possibility is that

Benefits from Higher Grading Standards Vary (Figure 3)

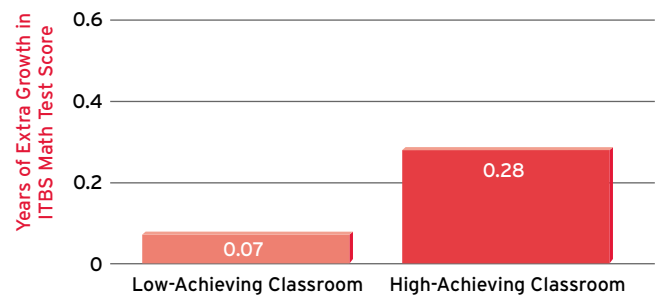
For high-achieving students, the positive effects of tough grading standards are larger when they are taught alongside relatively low-achieving peers. The opposite is true for low-achieving students.

Impact of Higher Grading Standards on High-Achieving Students*



* Student scoring half a standard deviation above the mean on ITBS reading test in the 3rd grade. Results significant at the .01 level.

Impact of Higher Grading Standards on Low-Achieving Students*



* Students scoring half a standard deviation below the mean on ITBS math test in the 3rd grade. Results significant at the .1 level.

Note: Graphs present the estimated impact of increasing teachers' grading standards by one standard deviation. The average test scores of classrooms labeled "low achieving" were 1.5 standard deviations below the mean for all classrooms. The average test scores in high-achieving classrooms were 1.5 standard deviations above the mean.

SOURCE: Authors

parents may devote more attention to their children's schoolwork if their grades suggest that they are struggling, as they might with a tough-grading teacher.

To assess the latter possibility, in the spring of 2001 we conducted a survey of parents with students in both 4th and 5th grades in Alachua County. We asked the responsible parent to report on how much time he or she spends weekly



What might explain the positive effects of higher grading standards? One possibility is that parents may devote more attention to their children's schoolwork if their grades suggest that they are struggling.

helping each of the two children with their homework. This allowed us to control for factors, such as parental motivation, that might be common to both siblings in a household. We found that, holding constant the child's grade level, 3rd-grade test scores, and the average 3rd-grade test score in the child's class, parents spend more time helping the child with the tougher teacher with homework than they do helping the

Tough grading standards elicit the most improvement from low-achieving students when they are in classrooms with relatively high overall achievement.

sibling with the easier teacher. These results do not appear to be due to tougher teachers' assigning more homework; parental reports suggest that the typical tough teacher assigns just 10 percent more homework than the typical easy teacher.

Another intriguing finding from this survey is that parents do not perceive tougher teachers to be better teachers. We asked parents to grade their children's teachers from A

to F. While there is relatively little variation in these grades (in their own form of grade inflation, two-thirds of the parents gave their children's teachers A's), the results suggest that, if anything, parents view tough teachers less favorably than they view easier teachers. Parents were 50 percent more likely to assign a grade of B or below to a tough teacher than to a relatively easy teacher, after adjusting for the same controls as above. This result suggests that our measure of grading standards is not merely reflecting some other attribute of a teacher that is viewed as desirable to parents. It also bolsters our argument that it is high grading standards rather than some unobserved measure of teacher quality that is responsible for the positive effects on students' performance gains.

Conclusion

Our results indicate that students benefit academically from higher grading standards. However, these results were not uniform: high-ability students appear to benefit more

than low-ability students from high grading standards. Moreover, initially low-performing students appear to benefit more from high grading standards when they are placed in high-achieving classrooms. Likewise, high-performing students appear to react best to high grading standards when placed in low-achieving classrooms.

It is, however, premature to conclude that high grading standards are unambiguously desirable. We cannot yet speak to the consequences of teacher-level grading standards at the secondary level, where the same effects might not hold and where high grading standards might even lead more students to drop out of school altogether. In addition, our results do not tell us anything about how to raise the grading standards of teachers whose standards are currently low. Before we can recommend a general policy of higher standards, we must understand the distributional consequences at all levels and know how to implement a policy of high standards. As the yawning gaps between the grades teachers presently give out and their students' scores on state tests suggest, it will be no easy task to change teachers' grading habits.

—David N. Figlio is a professor of economics at the University of Florida and a research associate of the National Bureau of Economic Research. Maurice E. Lucas is director of research and assessment for the school board of Alachua County, Florida.