

HIGH GRADING STANDARDS IMPROVE STUDENT PERFORMANCE

DAVID N. FIGLIO AND
MAURICE E. LUCAS *

Grading standards at universities throughout the world, even at the most prestigious institutions, have fallen dramatically over the last few decades, leading many academics and policy-makers to question whether students are learning as much at university as did their predecessors in prior generations. However, grading standards in primary and secondary education have received remarkably less attention. Very little discussion has been made in the policy arena on this topic, and even less attention has been paid to grading standards in the scholarly literature.

There are two major questions related to the analysis of grading standards in primary and secondary education. First, to what degree do the grades distributed by schools and teachers correspond to their students' performance on some objective measure of student performance, such as state and national exams? Second, and more important, how does "tough" or "easy" grading affect students' test performance and learning?

The literature on these questions is extremely thin. In fact, to our knowledge, the analysis presented here, describing our research earlier this year (Figlio and Lucas 2004), 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 and teacher-

to-teacher across three school years. Thus we can see how individual students perform on nationally norm-referenced exams as they move from "tough" to "easy" grading teachers and vice versa. Our results suggest that elementary students learn more with "tough" teachers, and that the magnitude of these effects varies 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 Alachua County, Florida school district, where Gainesville is located. This school district is relatively large by American standards, with about 1,800 test-taking students per grade, per year, and contains a wide variety of school settings, from urban to suburban to rural. Alachua County is racially heterogeneous, 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 (which in the United States is an indicator of low family income, as students in families below 185 percent of the national poverty line for their family size are eligible for subsidized lunches), while 19 percent are identified as gifted, 8 percent are learning disabled, and less than 1 percent are considered English learners. Our 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.

Florida provides a unique advantage for a study of this nature because it administers both the Iowa Test of Basic Skills (ITBS), a nationally normed test, and the Florida Comprehensive Assessment Test (FCAT). The ITBS can be thought of as an independent measure of how much students have learned in a school year along a generic standard that can be compared across a wide variety of educational systems, much like TIMSS and PISA are intended to do. The ITBS, however, like TIMSS



* David N. Figlio is Professor of Economics at the University of Florida and Research Associate at the National Bureau of Economic Research; Maurice E. Lucas is Director of Research and Assessment for the School Board of Alachua County, Florida.

and PISA, does not reflect the specific standards of the local educational system. But the FCAT was designed to measure the degree to which students are meeting the Sunshine State Standards, the same standards that are intended to be the basis for students' letter grades and promotion to the next grade. Students receive scores on the FCAT from 5 (highest) to 1 (lowest), with the thresholds for each performance level designed to correspond with the letter grades A (highest) through F (lowest). Thus, results from the FCAT are ideal for developing a measure of how generous teachers' grading policies are, and the ITBS is a useful measure for independently measuring how much students have learned according to another objective measure.

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. (We actually developed three different measures of grading standards, but the measure presented here elicits the most conservative results.) During the time of our study, students took 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 compare their gains in reading from 3rd to 4th grade, and in math from 4th to 5th grade, with their teachers' grading standards that academic year.

Grade inflation

On average, teachers tend to grade less stringently than the state standards (as reflected in FCAT scores) indicate that they should. For instance, just 9 percent of students who were awarded the highest letter grades by their teachers 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) as "proficient" on the

Florida standards, along with 61 percent of "B students" and 17 percent of "A students." Yet an important story is how different teachers are in their grading standards, not just across schools but also within schools, and indeed among colleagues at the same grade level in the same school. Even a crude example illustrates the stark difference across teachers in their grading standards: Among tougher-than-average teachers, 65 percent of A students attained level 4 or above while just 5 percent attained level 2 or below. For easier-than-average teachers, a mere 28 percent of students attained level 4 or above, while a remarkable 32 percent failed to make even the minimum standard for competency. Of course, looking at the quartile of easiest-grading teachers would provide an even more stark portrait of lax grading.

In short, teachers vary considerably in their grading standards, even within a single school district. And it turns out that teachers' 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. This finding is reassuring, since our empirical strategy relies mainly on within-school variation in teachers' grading standards to isolate the effects of those standards.

Teachers do not change their standards over time

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 actually a reflection of some other characteristic that might account for any effects we observe. Fortunately, from the researcher's perspective, 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 one year. This trend was essentially the same across the three categories, with very little movement between categories.

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 third grade test scores, in 1998–99 than the class they taught 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 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 none of these differences are statistically different. Tough graders are more likely to hold master’s degrees, a difference that is statistically significant. 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 may drive the estimated effects of grading standards on student outcomes.

Classroom assignment

The method by which students are assigned to teachers can also cause problems for the researcher. The fact that students may not be randomly assigned to teachers would be especially troublesome in a cross-sectional analysis, in which one compares one classroom to 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 to be low-income or learning disabled. This is true even within a school. Hence, it is unclear whether the outcomes associated with high standards are actually due to the standards themselves, or to some factor that is associated both with high-achieving students and the teachers to whom they tend to be assigned.

But our analysis looks at year-to-year changes in the grading standards faced by a given 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, 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 subgroups—black and white students are approximately equally likely to transition between groups, as are free-lunch-eligible and ineligible students. It is the case that gifted students, 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. Nor do the results presented below change materially when gifted students are excluded from the analysis.

Empirical results

We performed multiple analyses, progressively adding controls for students’ and teachers’ characteristics and the characteristics of their classrooms and schools as we went along. For our primary analysis, we controlled for the average annual gain made by all students in the relevant school during the period of analysis, such classroom characteristics as the share of white students, the share eligible for free lunches, and the students’ average math score in 3rd grade, as well as 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.

Nearly all of our analyses found statistically significant relationships between higher standards and improved performance and behavioral outcomes, though the magnitude of the improvement differed depending on which characteristics were controlled for. Our primary analysis, which includes all of the control variables of interest, found modest, statisti-

cally significant improvements in test scores associated with higher standards, and modest improvements in behavior that were not statistically significant. We found that moving from a low-standards teacher to a high-standards teacher was associated with over a half year's worth of gains in reading and mathematics—the same type of test score difference typically seen with regard to the difference between a child of college-educated parents vis-à-vis a child of a high-school dropout, and a larger effect than has been found associated with very large reductions in class sizes or very large increases in teacher salaries. (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 there may be substantial differences in how students experience standards, with higher standards producing both winners and losers. For instance, those 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. In their empirical study of grading standards in secondary school Betts and Grogger (2003) found that high-performing students benefited the most from high grading standards.

To study this issue, we tested whether the effect of high grading standards differed for students with different initial test scores. We found that an average student in 3rd grade benefits strongly (and significantly) from higher grading standards, with above-average initial performers benefiting as well. In addition, the results suggest that higher grading standards exert a significantly positive influence on students who are no more than 0.8 and 0.9 standard deviations below the average score in reading and math, respectively. However, the estimated effects of grading standards are negative for less than 1 percent of the student population, and never statistically significant.

We also examined how a classroom's overall achievement level, as measured by their average

third grade test score in the relevant subject, interacts with grading standards. We found that higher-achieving classes may fare somewhat better than lower-achieving classes under teachers with tough grading standards.

What may be more interesting, however, than the performance of entire classes is the distributional effect *within* a class. 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 of relatively low overall achievement. 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.

For instance, a student whose score in 3rd grade was half a standard deviation below the mean experienced nearly a third of a year of extra growth in learning when a teachers' degree of toughness was raised by one standard deviation in a classroom with high overall achievement (where the average score is 1.5 standard deviations above the mean). This compares with an improvement of just 0.07 of a year in a high-standards classroom with relatively low achievement. Similarly, for a student whose 3rd-grade reading performance is 1.5 standard deviations above the mean, the estimated effect of increasing teacher toughness by one standard deviation ranges from 0.18 years of extra growth (in a classroom averaging 1.5 standard deviations above the mean) to 0.71 years (in a classroom averaging 1.5 standard deviations below the mean).

This result has intuitive appeal. Given that the distribution of grades within a class varies much less across classes than does the distribution of performance on external assessments, one can assume that high-achieving students are more likely to earn high grades in low-achieving classes than they are in high-achieving classes. Likewise, low-achieving students in high-achieving classes are at relatively more risk of receiving a low grade than are low-achieving students in low-achieving classes. Hence, it seems sensible that initially high-achieving students are challenged more to get a "good grade" with tough teachers, particularly 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 teach-

ers, 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 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 spring 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 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 systematically spend more time helping the child with the tougher teacher with homework than they do helping the sibling with the easier teacher. Indeed, we estimated that a parent of a child with a teacher with tougher grading standards than 75 percent of all teachers would spend 60 percent more time helping that child than he or she would spend with that child's sibling who had a teacher with grading standards tougher than only 25 percent of teachers.

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. This is consistent with findings from interviews with principals in the district, who reported that teachers within any given grade level in the school work to assign the same amount of homework per week. Unfortunately, we have no way of judging whether the homework assigned by tougher teachers is more challenging than that assigned by easier-grading teachers, but casual evidence suggests that it tends to be very similar in nature and difficulty.

Another interesting finding from this survey is that parents do not perceive tougher teachers to be bet-

ter teachers. We asked parents to grade their children's teachers from A to F. While there is relatively low 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, both in their test-score performance and on measures of behavior in school. 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-ability classrooms. Likewise, high-performing students appear to react best to high grading standards when placed in low-ability classrooms.

It is, however, premature to conclude from this study that high grading standards are unambiguously desirable. We cannot yet speak to the distributional consequences of teacher-level grading standards at the secondary grades, where it may be the case that high grading standards would lead more students to drop out of school altogether. Lillard and DeCicca (2001) have found, for instance, that raising *graduation* standards tends to lead to higher rates of student drop-outs as well. In addition, the results here 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, it is important to understand the distributional consequences at all levels, as well as to know how to implement a policy of high standards.

References

Betts, J. R. and J. Grogger (2003), "The Impact of Grading Standards on Student Achievement, Educational Attainment, and Entry-Level Earnings", *Economics of Education Review* 22(4), August, 343-52.

Figlio, D. N. and M. E. Lucas (2004), "Do High Grading Standards Affect Student Performance?", *Journal of Public Economics* 88(9), 1815-34.

Lillard, D. R. and P. DeCicca (2001), "Higher Standards, More Drop-outs? Evidence Within and Across Time". *Economic of Education Review* 20(5), 459-73.