

Where A Is Ordinary: The Evolution of American College and University Grading, 1940-2009

STUART ROJSTACZER

Independent Scholar

CHRISTOPHER HEALY

Furman University

Background/Context: *College grades can influence a student's graduation prospects, academic motivation, postgraduate job choice, professional and graduate school selection, and access to loans and scholarships. Despite the importance of grades, national trends in grading practices have not been examined in over a decade, and there has been a limited effort to examine the historical evolution of college grading.*

Purpose/Objective/Research Question/Focus of Study: *Here we look at the evolution of grading over time and space at American colleges and universities over the last 70 years. Our data provide a means to examine how instructors' assessments of excellence, mediocrity, and failure have changed in higher education.*

Data Collection and Analysis: *We have collected historical and contemporary data on A–F letter grades awarded from over 200 four-year colleges and universities. Our contemporary data on grades come from 135 schools, with a total enrollment of 1.5 million students.*

Research Design: *Through the use of averages over time and space as well as regression models, we examine how grading has changed temporally and how grading is a function of school selectivity, school type, and geographic region.*

Findings/Results: *Contemporary data indicate that, on average across a wide range of schools, A's represent 43% of all letter grades, an increase of 28 percentage points since 1960 and 12 percentage points since 1988. D's and F's total typically less than 10% of all letter grades. Private colleges and universities give, on average, significantly more A's and B's combined than public institutions with equal student selectivity. Southern schools grade*

more harshly than those in other regions, and science and engineering-focused schools grade more stringently than those emphasizing the liberal arts. At schools with modest selectivity, grading is as generous as it was in the mid-1980s at highly selective schools. These prestigious schools have, in turn, continued to ramp up their grades. It is likely that at many selective and highly selective schools, undergraduate GPAs are now so saturated at the high end that they have little use as a motivator of students and as an evaluation tool for graduate and professional schools and employers.

Conclusions/Recommendations: *As a result of instructors gradually lowering their standards, A has become the most common grade on American college campuses. Without regulation, or at least strong grading guidelines, grades at American institutions of higher learning likely will continue to have less and less meaning.*

Unregulated and self-regulated organizations and professions commonly have problems maintaining standards and ensuring ethical behavior (DeMarzo, Fishman, & Hagerty, 2005; Frey, 2006). Without regulation, it is likely that many individuals will pursue local benefits even if their actions are detrimental to the global good. Grading of undergraduates at American colleges and universities incorporates a system of standards that is almost always unregulated. The A–F letter grade system of American higher education has been in wide use for roughly the last 100 years (e.g., Meyer, 1908) and gradually became the basis for the now ubiquitous 4.0 grade point average (GPA) scale. Implicit in the use of our grading system is the belief that it has value both as a motivator of students and as a tool for postgraduate schools and employers to identify the best and brightest. The assumption is that college instructors will individually regulate their grading practices out of a sense of personal integrity and because they realize that how they grade and teach influences the reputations of the institutions they represent.

Efforts at employing even soft external guidelines on grading practices are almost always rebuffed by both instructors and university leadership, and often equated with a lack of faith in the integrity of the faculty. A grade should reflect an instructor's true view of student performance, but a college instructor may be at best ambivalent about the worth of grades (e.g., Battersby, 1973). Even if an instructor feels that grades have value and purpose, there are significant perceived incentives for that instructor to abandon any objective standard and award grades that are artificially high (e.g., Feldman, 1976; Johnson, 2003). In the absence of oversight, one might expect that grading standards at colleges and universities would degrade over time. A's would become commonplace. Failing and substandard grades would become rare. The ability of grades to motivate or serve as an indicator of performance would be impaired.

Have universities and colleges managed to maintain academic standards in the absence of regulation? We have tried to answer that question by examining undergraduate grade distributions over the last 70 years using historical and contemporary data from over 200 American four-year schools (institutions are listed in the appendix). We measure changes in academic standards over time, examine the evolution of the divergence in grading practices between public and private schools, and look at the potential causes of those changes.

METHODOLOGY

We assembled our data on four-year school grades (grades given in terms of percent A–F for a given semester or academic year) from a variety of sources: books, research articles, random World Wide Web searching of college and university registrar and institutional research office Web sites, personal contacts with school administrators and leaders, and cold solicitations for data from 100 registrar and institutional research offices, selected randomly (20 of the institutions solicited agreed to provide contemporary data as long as the school’s grading practices would not be individually identified in our work).

The characteristics of the 135 institutions for which we have contemporary data are summarized in Table 1. In addition, we have historical data on grading practices from the 1930s onward for 173 institutions (93 of which also have contemporary data). Time series were constructed beginning in 1960 by averaging data from all institutions on an annual basis. For the 1930s, 1940s, and 1950s, data are sparse, so we averaged over 1936 to 1945 (data from 37 schools) and 1946 to 1955 (data from 13 schools) to estimate average grades in 1940 and 1950, respectively. For the early part of the 1960s, there are 11–13 schools represented by our annual averages. By the early part of the 1970s, the data become more plentiful, and 29–30 schools are averaged. Data quantity increases dramatically by the early 2000s with 82–83 schools included in our data set. Because our time series do not include the same schools every year, we smooth our annual estimates with a moving centered three-year average. It is worth noting that the same trends we detail here are also clearly visible in the unsmoothed data. They are also clearly visible if we reduce our database to the 14 schools for which we have mostly continuous data from the 1960s (or earlier) to the 2000s.

Table 1. Characteristics of Schools With Contemporary Data Including Grading Averages

Classification	Number of Schools	Total Population	Avg. SAT 2006	Avg. SAT 1966	Avg. %A	Avg. %B	Avg. %C	Avg. %D	Avg. %F
<i>By school type</i>									
Private nonprofit university	9	46,788	1245	1235	48.2	35.8	11.4	2.2	2.3
Private nonprofit college	34	75,348	1192	1184	47.7	36.6	11.3	2.4	1.9
Public flagship university	26	532,034	1172	1139	42.3	34.5	15.5	4.1	3.6
Public satellite university	25	282,449	1056	1050	41.7	32.0	16.0	4.8	5.4
Public commuter university	41	579,070	1017	1018	39.0	31.8	17.5	5.4	6.3
<i>By region</i>									
Midwest	28	281,092	1135	1147	45.0	34.0	14.0	3.7	3.3
Northeast	28	208,209	1153	1184	45.1	35.7	13.0	3.1	3.1
West	29	377,936	1079	1094	44.6	33.0	14.4	3.7	4.2
South	50	648,452	1102	1055	39.7	33.1	16.7	5.1	5.4
Totals	135	1,515,689	1115	1111	43.0	33.8	14.9	4.1	4.2

To examine controls on grading, we employed three ordinary least squares regression models. In our first model, contemporary data were separated by school type (private vs. public) to examine the relationship between SAT scores of students and the sum of the percentage of A and B grades at the schools in our database. We chose the median SAT score of a school's matriculates as a parameter because it is a good surrogate for institutional selectivity and because earlier work indicates that institutionally averaged SAT scores are well correlated with institutionally averaged GPAs (Astin, 1971; Rojstaczer & Healy, 2010). Our first model examined the following simple relationship between grades and SAT scores in our contemporary database for private schools and public schools separately:

$$\text{Percent (A+B)} = a_1 + a_2\text{SAT} \quad (1)$$

where a_1 represents the intercept, a_2 represents the estimated linear coefficient between SAT scores and grades, and SAT is the sum of the math and verbal components of the median SAT score of matriculates at a school (2006 data; National Center for Education Statistics [NCES], 2009a).

Our second model examined the sum of the percent A and B grades as

a function of 2006 SAT scores, time, region, school type (private or public), and school focus over the period 1988–2007:

$$\text{Percent (A+B)} = a_1 + a_2\text{SAT} + a_3\text{YEAR} + a_4\text{SOUTH} + a_5\text{PRIVATE} + a_6\text{STEM} \quad (2)$$

where a_1 and a_2 are the same as in Equation 1 (but determined independently with Equation 2), a_3 is the estimated linear coefficient between time and grades, YEAR is the academic year since 1988, a_4 is the estimated difference between grades in the South versus the rest of the United States, SOUTH is an indicator variable (1 if the institution is from the South, 0 if it is not), a_5 is the estimated difference between grades at private schools and public schools, PRIVATE is an indicator variable (1 if the institution is private, 0 if it is not), a_6 is the estimated difference between grades at schools with a science and engineering focus (schools that are known mostly for science and engineering or the flagship state school where science and engineering are emphasized) and those more liberal arts oriented, and STEM is an indicator variable (1 if the institution has a science and engineering focus, 0 if it doesn't). We included these indicator variables because the data suggested that in addition to grades being a strong function of school selectivity and time, both Southern schools and engineering schools had low grades, and private schools had high grades in comparison with national averages (Table 1).

Our third model examined the percent A grades as a function of 2006 SAT scores, time, region, school type (private or public), and school focus over the period 1988–2007:

$$\text{Percent A} = a_1 + a_2\text{SAT} + a_3\text{YEAR} + a_4\text{SOUTH} + a_5\text{PRIVATE} + a_6\text{STEM} \quad (3)$$

where the coefficients are the same as in Equation 2, recalibrated for a different dependent variable.

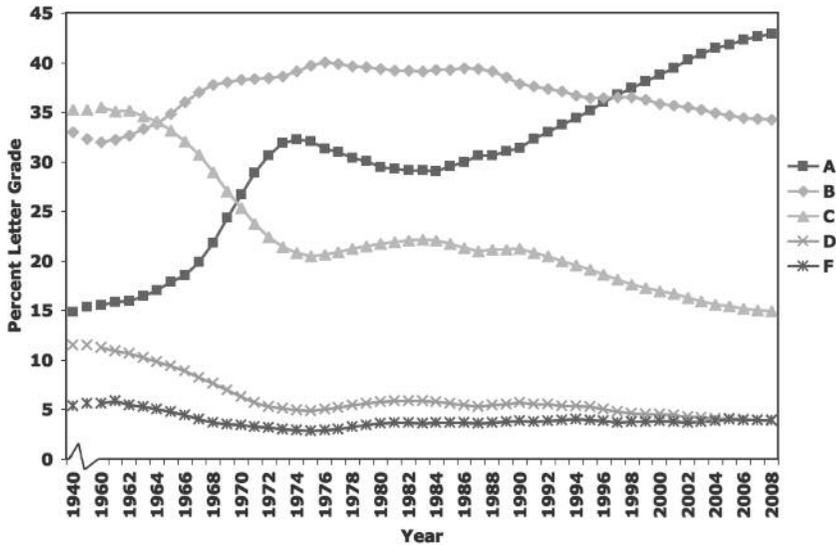
RESULTS

Our examination of nationwide trends indicates that grading practices were largely constant for decades, but grade distributions have undergone gradual yet very significant changes since the 1960s. For the schools in our database, the number of A's awarded has increased to such a degree that A is now ordinary. On average, A is now by far the most common grade awarded on American four-year campuses. Substandard grades, D and F, typically are awarded less than 10% of the time even on campuses with students of modest academic caliber.

Our data (Figure 1) show that in 1960, as in the 1940s and 1950s, C was

the most common grade nationwide; D's and F's accounted for more grades combined than did A's. On average, instructors were assigning grades by using a slightly skewed normal distribution curve centered at about a C+. By 1965, however, B had supplanted C as the most common grade, and D's and F's were becoming increasingly less common. From the early 1960s to the mid-1970s, grades rose rapidly across the nation, and A became the second most common grade awarded.

Figure 1. Distribution of grades at American colleges and universities as a function of time



Note: 1940 and 1950 (nonconnected data points in figure) represent averages from 1935 to 1944 and 1945 to 1954, respectively. Data from 1960 onward represent annual averages in our database, smoothed with a 3-year centered moving average.

The Vietnam era was followed by a decline in A's that lasted for roughly a decade. Awarding of A's began to rise again in the mid-1980s. From 1984 to the mid-2000s, the proportion of A's increased by a factor of 1.5. By 2008, A's were nearly three times more common than they were in 1960.

Our data on historical grade distribution averages agree well with other studies that have compiled grade distributions and GPAs from university-based data (Edson, 1955; Juola, 1976, 1980; Perry, 1943; Suslow, 1976). Indeed, we decided to incorporate the grade distribution data of older work to supplement our own database (Perry; Suslow). They also agree

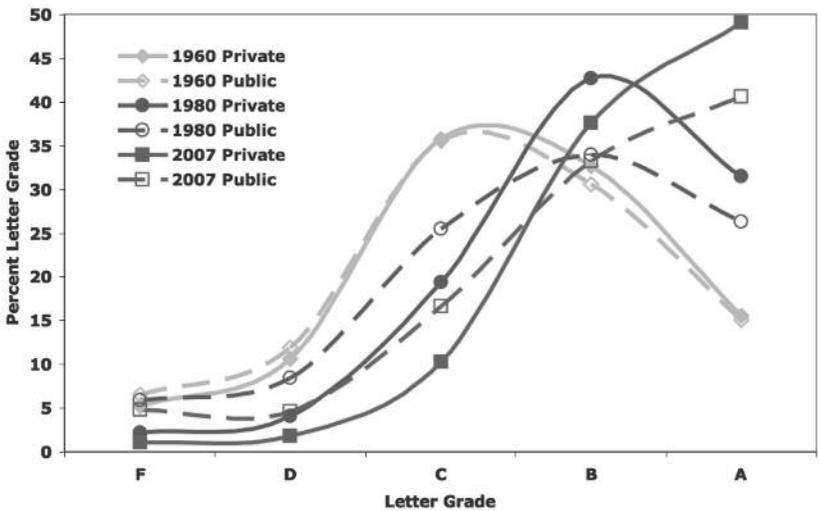
with historical transcript-based data for nonselective and selective colleges (Adelman, 2004). Our data do not match well with earlier work on transcript-based data for highly selective colleges (Adelman, 2004), but this is not surprising given that those earlier estimates used fewer than 900 sets of transcripts from highly selective colleges nationwide over roughly a 27-year interval. Historical self-reported student data also suggest significant rises in GPAs (Kuh & Hu, 1999; Levine & Cureton, 1998), but the changes they show are larger than those found here and in other studies; the differences likely point to the difficulty of using self-reported student data in quantitative assessments. Given that our historical data agree with other studies that have extensively compiled grades from transcripts and university-based data from the mid-1990s and earlier, we are confident that our post-mid-1990s data also represent national trends.

Averages mask variations, and it is worth looking at recent interinstitutional variability in both grade inflation and contemporary grading patterns. For the 12-year period from 1997 through 2008, we have extensive (7 years or more in length) data from 79 colleges and universities. Fourteen of those schools have modest to negative grade inflation (a less than 2 percentage point increase in A's) over 1997–2008. At three of those schools—Central Michigan, Colorado, and Princeton—leadership (locally at the school level or at the state legislative level) has made a concerted effort to influence grading. All but one of the other 11 colleges are public schools, and 6 of them award A's over 40% of the time; they appear to have reached a plateau, perhaps temporary, in A grades awarded. Another four are from the South, a region that tends to give lower grades (Table 1). The remaining school is in the California state system and has faced financial pressure to lower admissions standards; these poorer quality students apparently have caused grades to drop.

But for the remaining 82% of the schools for which we have extensive data over the period 1997–2008, grade rises are significant. Rates of inflation are much higher at private schools than they are at public schools, with the increase in A's averaging 9.6 and 5.2 percentage points, respectively. All schools in our database have seen significant rises in grades since the mid-1980s.

On a national basis, the evolution of grading practices seems to be the result of a gradual abandonment of curve-based grading (Figure 2). Grading practices for private and public schools, which were similar prior to the 1960s, were quite different by the 1980s. By the late 2000s, A's and B's represented 73% of all grades for public schools and 86% of all grades for private schools in our database.

Figure 2. National average grading curves as a function of time, 1960, 1980, and 2007 for public and private schools



Note: 1960 and 1980 data represent averages from 1959–1961 and 1979–1981, respectively.

It has been observed that colleges and universities nationwide have, since the post-Vietnam era, established a national ad hoc grading system in which the average GPA of an institution is dependent on the selectivity of that school's admissions (Astin, 1971; Rojstaczer & Healy, 2010). Not surprisingly, this trend is similarly present in our grade distribution data. Our simple regression model (Equation 1) indicates that the average percentage of A plus B grades at a school correlates significantly with the median SAT score of its student body (Figure 3), with private schools, on average, awarding more high grades than public schools.

For the 135 schools in our database with contemporary data, A's are handed out 43% of the time on average (42% when weighted by total student population rather than the simple average of all schools; Table 1). As can be expected in a grading system that is dependent on the average student quality of a school and school type (private vs. public), there is considerable variation about that mean value on a school-by-school basis (Figure 4).

Figure 3. Relationships determined by linear regression between the median SAT score (for 2006) of matriculates at a school and the percentage of A plus B grades awarded in our contemporary database for schools with SAT data (130 schools with a population of 1.5 million undergraduates)

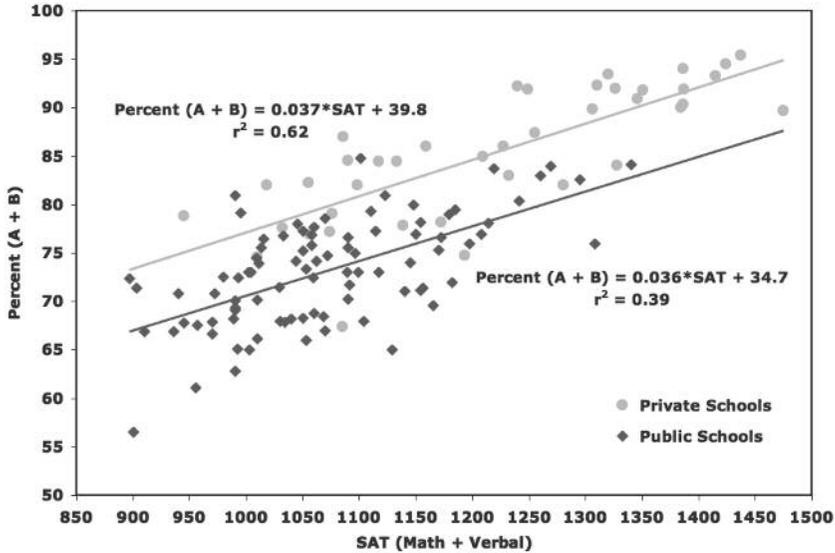
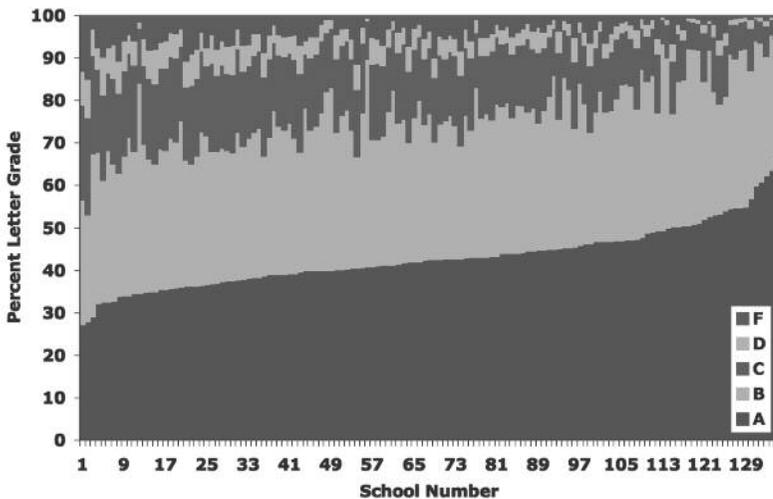


Figure 4. Most recently available average undergraduate grade distribution curves for the 135 schools with contemporary data in our database



Note: 90% of all data come from 2007, 2008, or 2009. Schools are numbered in order of percent A's awarded.

Our sample has a student population of 1.5 million, far greater than any other previous detailed study on national grading patterns for four-year colleges and universities. It should be noted, however, that although we randomly found and sought data, in comparison with national student populations, our sample underrepresents private schools (which grade higher than national averages) and overrepresents Southern schools (which grade lower than national averages) (Aud et al., 2011; NCES, 2008a). These differences between our data set and national enrollment patterns partly reflect the reticence of private schools to divulge grading data. Southern schools, in comparison with schools in other regions, far more commonly post their grading data on the World Wide Web; they were also slightly more likely to respond positively to our cold solicitations for data. The average SAT score of our sampled student body weighted by student population (math plus verbal) is about 40 points higher than that seen nationally for 2008 in a survey of 2,343 four-year institutions (1105 vs. 1064; NCES, 2009b).

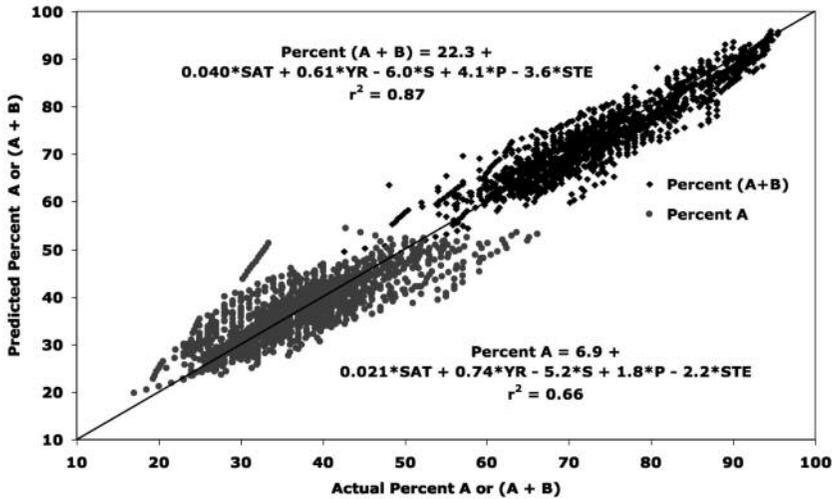
As a check on whether biases in our sampling significantly distort our average estimates, we examined the deviation in our sample from an expected normal distribution. Our sample set displays moderate skewness (0.70) and kurtosis (0.90), and a chi-square test indicates that we cannot reject the hypothesis that our sample distribution is Gaussian (two tailed p value of 0.23) (Spiegel & Stephens, 1998).

The combined effect of undersampling private schools, oversampling Southern schools, and (probably) the slightly higher average SAT scores of our sampled students relative to national averages suggests that our weighted average of 42% A's is a slightly conservative one. Our regression model of Equation 1, which does not correct for our Southern school bias but can correct for our slightly higher average SAT scores, yields the same estimate for percent A grades nationally.

Our regression models of Equation 2 and Equation 3 (Figure 5) further reinforce the significance of our estimates of national grading trends based on simple and weighted averages of our sample over time and space. As with our regression model of Equation 1 (Figure 3), grades are strongly dependent on SAT scores, and the SAT regression coefficient determined independently with Equation 2 is essentially the same as that determined with Equation 1. Comparison of the SAT coefficient estimated from Equations 1 and 2 with that estimated for percent A's only (Equation 3) suggests that at highly selective schools, both A's and B's are significantly more common than they are at less selective schools.

The long-term trend of increasing A grades inferred from the regression models of Equation 2 and 3 (a 7–8 percentage point increase in A's per decade) is slightly greater than that determined from simply averag-

Figure 5. Predictive regression models of the percentage of A or A plus B grades awarded as a function of the median SAT score (for 2006) of matriculates at a school, time since 1988 (YR), whether the school is in the South (S), whether the school is private (P), and whether the school has a science and engineering focus (STE) in comparison with observed grades over the time period 1988–2007 (1,371 data points for 119 schools)



ing all data over time. Similar to our regression model of Equation 1, multiple regression modeling suggests that private schools award about 5% (4 percentage points) more A and B grades combined than public schools of equal selectivity; B grade increases perhaps are slightly more significant than A grade increases. Southern schools grade about 6 percentage points lower (A and B grades combined) than elsewhere in the country, and our regression modeling results are consistent with our simple averages (Table 1) in that they suggest that the difference is dominated by fewer A grades. Science and engineering-focused schools are less likely to award both A and B grades in comparison with other schools of equal selectivity.

The estimated regression coefficients of Equations 2 and 3 (Figure 5) have standard errors of 10% or less and 20% or less of their absolute values, respectively. Both multiple regression models suggest that nationwide as of 2007—given an average nationwide SAT score of 1064—43% of all letter grades were A on average. In summary, all our estimates, regardless of whether we adjust for sample bias, suggest that during 2007–2008, A’s represented about 43% of all letter grades earned at American four-year colleges and universities.

DISCUSSION

TEMPORAL TRENDS AND INSTITUTIONAL VARIABILITY

The rise of GPAs in the 1960s was identified long ago (Juola, 1976, 1980), although actual changes in letter grades awarded were not identified explicitly. Our data are consistent with the conventional wisdom that 1960s and early 1970s grade inflation was caused by instructors abandoning D's and F's so that students could avoid the Vietnam-era military draft. But C's also dropped precipitously, and A's rose dramatically. Professors, in their generous grading during the Vietnam era, were doing more than simply keeping students from being removed from school; increasingly, they were abandoning traditional bell-shaped, curve-based grading altogether and were grading as if students produced work that was almost always good to excellent in quality.

For about a decade, grades declined slightly after the Vietnam era. It would be difficult to ascribe this decline to national changes in student quality. Average SAT scores for college-bound high school seniors did decline about 30 points (math and verbal combined) over the early to late 1970s, but similar declines were observed during the Vietnam era of grade rises; also, SAT scores began to rebound modestly in 1981 (College Board, 2007). Women, who traditionally do better than men in college, also began to increasingly enter colleges and universities over this period.

We theorize that this era of slightly more stringent grading may have been caused by an ad hoc national response on the part of the professorate to the end of military draft pressures on the student body. Whatever its cause, C's, D's, and F's became slightly more common at the expense of a reduction in the percentage of A's awarded from the early 1970s to the early 1980s.

The resumption of grade inflation in the mid-1980s had yet to end as of 2008. Although A's increased dramatically over this roughly 25-year period, the percentage of D grades dropped only slightly, and the number of F grades held steady. Unlike in the 1960s, grade inflation was no longer raising all boats in the late 2000s. It was elevating the grades of the good and mediocre, but it was still possible, at least in public schools, for a significant number of students to fail. The character of grade inflation since the 1980s helps explain why contemporary students who drop out often continue to have very low GPAs (e.g., Stassen, 2001).

There is no indication that the rise in grades at public and private schools has been accompanied by an increase in student achievement. If anything, measures indicate that student performance has declined. Scholastic aptitude test scores of potential college applicants dropped to

such a degree that they had to be recentered upward in the 1990s (Dorans, 2002). Over the 1990–2008 period, average SAT scores of college-bound students (math plus verbal) have gone up by less than 20 points (College Board, 2007), and in our database, the rise is less than 10 points since the mid-1960s (Table 1). Students are studying about 10 hours a week less today than they did in the 1960s (Babcock & Marks, 2010). Literacy of graduates has declined (Kutner, Greenberg, & Baer, 2006). Student engagement levels are at all-time lows (Saenz & Barrera, 2007). Colleges and universities likely are awarding significantly higher grades despite national declines in student achievement.

It is worth noting that despite rising grades, graduation rates of students have remained largely static for decades (Kuh, Kinzie, Schuh, & Whitt, 2005). This is not a paradox. Many factors, aside from grades, influence graduation rates. These include the dramatic rise in the cost of higher education since the 1970s (e.g., Ehrenberg, 2002). Also, as noted earlier, many low achievers are still receiving poor grades.

Even if grades were to instantly and uniformly stop rising, colleges and universities are, as a result of five decades of mostly rising grades, already grading in a way that is well divorced from actual student performance, and not just in an average nationwide sense. A is the most common grade at 125 of the 135 schools for which we have data on recent (2006–2009) grades (Figure 4). At those schools, A's are more common than B's by an average of 10 percentage points. Colleges and universities are currently grading, on average, about the same way that Cornell, Duke, and Princeton graded in 1985. Essentially, the grades being given today assume that the academic performance of the average college student in America is the same as the performance of an Ivy League graduate of the 1980s.

CAUSE OF GRADE RISES: PREVIOUSLY INVOKED INFLUENCES THAT ARE LIKELY MINOR OR INSIGNIFICANT

What has caused grades to rise? Over the decades covered by our data, there have been tremendous changes in the nature of higher education, including a more than two-fold increase in enrollment in four-year colleges since 1960 (Aud et al., 2011; NCES, 2009c), a significant increase in female and minority students (King, 2000, 2010; NCES, 2008b), an increase in the percentage of students employed while attending school (NCES, 2005a), and changes in pedagogic approaches. As we examine next, none of these factors appears to be the main driver of changing patterns in grading. In the Vietnam era, there was little debate as to why grades were rising: Professors clearly were grading more generously.

Lowering of standards appears to be the principal reason that grades began to rise in the 1980s as well.

Some have made claims that rises in grades at their institutions are primarily the result of better students (e.g., Weinberg, 2007). At an institutional level, there have been local increases in student quality that are well documented and no doubt account for some of the observed increases in grades. This, however, is a secondary effect. For example, at New York University, a small increase in the average GPA of the graduating class of 2001 relative to the class of 1997 may have been strongly influenced by increases in student quality (Weinberg) over that short period, but high grades were already present in 1997 (60% of the class of 1997 had a B+ GPA or better; in contrast, only 10% had a GPA less than B-). Similarly, at Penn State, average SAT scores increased by 70 points (math and verbal combined) from 1966 to 2006, which we estimate, from regression of SAT scores versus grades in our database (see Figures 3 and 5; similar relationships between GPAs and SAT scores can be found in McSpirit & Jones, 1999; Rojstaczer & Healy, 2010; and Vars & Bowen, 1998), should correspond to about a 1–2 percentage point increase in awarded A's. Over that 40-year period, A's increased by 21 percentage points at Penn State, and more than half of that rise took place after 1983.

Another secondary influence related to better student performance may be the rise of women as a percentage of total national undergraduate enrollment. Our data indicate that women generally perform better than men, typically 0.1 to 0.2 better on a 4.0 GPA scale. But gender ratios stabilized in the 2000s after a rise from about 50/50 to 57/43 from the 1970s through the 1990s (King, 2000, 2010; NCES, 2008b). As noted earlier, much of that rise took place during a period when, on average, grades fell slightly.

One could also argue that grades are rising as a result of gradual improvements in pedagogy over time. There are no data to show that this is the case, although we are inclined to believe based on personal experience that some newer teaching approaches have indeed improved student performance. Yet pedagogic approaches weren't static prior to the era of grade inflation, nor were they static over the mid-1970s to mid-1980s. It would be far-fetched to claim that teaching has improved to such a degree that students are performing dramatically better—and have made excellence the most common level of achievement—while working less and being less engaged in studies overall.

Other attempts to find reasons for observed grade rises have focused on student course choice and enrollment preferences. For example, some have ascribed grade rises to students gaming college grading.

Students may be increasingly using new online course-grading databases at places like Cornell (Bar, Kadiyali, & Zussman, 2008) and professor-rating databases nationwide (Kaplan Test Prep and Admissions, 2010) to find easy courses and professors. But our examination of Cornell data indicates that A grades were increasing at that university at the same rate before the creation of student-accessible grading databases as after. Our national data also indicate that grade rises did not swing upward noticeably with the increased use of professor-rating databases in the mid-2000s; in fact, rates of grade rises may have decreased slightly.

There have also been attempts to ascribe at least some of the change in grade distributions to increases in nontraditional grades (e.g., Adelman, 2004). These changes were significant in the 1960s and 1970s. However, Pass, No Credit, Withdraw, and other non-letter grades in our database over the period of recent grade inflation (1984–2009) do not commonly change by more than a few percent. Because of a post-1980s student shift away from the Pass/No Credit grading option that became popular in the 1960s and 1970s, decreases in non-letter grade percentages are more common than increases.

Changes in choice of student majors may have influenced average grades as well. But these student choice changes may have caused a slight retardation in grade inflation. From 1970 to 2003, the percentage of students graduating in the sciences, an area of study with traditionally more stringent grading (Rojstaczer & Healy, 2010), was essentially unchanged if one includes the field of computer science (NCES, 2005b). On the other hand, the percentage of majors in the humanities, where grading tends to be more generous (Rojstaczer & Healy), decreased dramatically (Chace, 2009). It seems unlikely that course choice—whether it has involved increased gaming to preferentially select easy classes, or withdrawing early, or choosing a Pass/Fail option, or choice of major—has had a significant influence on grading patterns over the last 25 years.

Some on the political right in America have noticed the rise in college grades and have attempted to ascribe it to the behavior of left-leaning, permissive faculty. Some have also made the claim that affirmative action and the rise of minority student enrollments have caused grade inflation (e.g., Mansfield, 2001) even though much of the increase in minorities relative to national demographic trends, like the increase in women, took place during a period of declining grades (NCES, 2008b). Although culture does play a role in grading trends—as is evident in the differences in grading between private and public schools, as well as the relatively low grades in Southern and science and engineering-focused schools—we view grade inflation and the abandonment of realistic student evaluation over the last 25 years to be mostly independent of political leanings.

Finally, there have been studies and speculation about the role of adjuncts and non-tenure-track faculty in rising grades (e.g., Kezim, Pariseau, & Quinn, 2005). But at most of the schools in our database that grant a high percentage of A's, tenure-track faculty continue to play the dominant role in teaching. In contrast, some of the commuter colleges in our database, which have had significant changes in full-time to part-time faculty ratios, have some of the lowest grades. Rising grades are the result of the actions of both tenure-track and non-tenure-track instructors.

INCENTIVES FOR GRADING MORE GENEROUSLY

Given the absence of other significant factors on rising grades, it is likely that the major cause of grade inflation over the last 25 years is the same as the cause of grade inflation in the Vietnam era: Instructors at American colleges and universities have been grading more generously. In the absence of oversight, and because of the presence of positive incentives to give artificially high grades, higher education has gradually abandoned its grading standards.

In the 1960s and early 1970s, the incentives for professors to grade generously were external and political. The absence of those external incentives may have been the reason that grades declined slightly from the mid-1970s to the early 1980s. What incentives were in place for grade inflation to return in the 1980s? This decade coincided with the establishment of a new approach toward students by leaders in higher education: Students were no longer considered acolytes, but consumers of a product (e.g., Bayer, 1996; Zirkel, 1994). Describing the reasons for this change is beyond the scope of this article, but the new approach significantly changed the ethos of college campuses. A former chancellor from the University of Wisconsin–Madison has described well the national transition in campus culture and teaching:

That philosophy (the old approach to teaching) is no longer acceptable to the public or faculty or anyone else. . . . Today, our attitude is we do our screening of students at the time of admission. Once students have been admitted, we have said to them, "You have what it takes to succeed." Then it's our job to help them succeed. (Finkelmeier, 2010)

The establishment of a consumer-based approach to teaching has created both external and internal incentives for the faculty to grade more generously. Externally, higher grades translate into better future prospects for graduates. This motivation was stated explicitly as the reason that Haverford changed its grading policies campuswide in the early 1970s (Faculty of Haverford College, 1972), and we've found evidence in

printed faculty governance meeting discussions for a desire to keep grades high elsewhere. Faculty members may sense that institutions similar to their own are raising grades and feel a need to keep up so as not to disadvantage their future alumni (e.g., Perrin, 1998). Scholarship programs that demand minimum GPAs may also provide local external incentives (e.g., Georgia Student Finance Commission, 2011).

The consumer-based approach to undergraduate education has also resulted in a desire to keep students pleased with their class experiences and to measure the degree to which students are satisfied with their instructors and classes. Student-based course evaluations rose in prominence on college campuses in the 1970s, and their use became increasingly common through the 1980s and 1990s. In 1973, 26% of 600 surveyed liberal arts colleges employed student ratings of teachers; in 1983 and 1993, the numbers rose to 68% and 86%, respectively (Seldin, 1984, 1993). These evaluations are used for more than just improving instruction; in the 1980s and 1990s, they increasingly began to play a minor to significant role in decisions regarding pay, promotion, and retention of instructors. The widespread use of course evaluations likely has had a profound influence on grading.

There has been extensive work on the relationship between course evaluations and grades, and much evidence indicates that higher grades yield more positive student-based evaluations (e.g., Ellis, Burke, Lomire, & McCormack, 2003; Feldman, 1976; Johnson, 2003; Weinberg, Fleisher, & Hashimoto, 2007), although some have disputed that evidence (e.g., Marsch & Roche, 1997). Regardless of whether such a relationship is strong, weak, or nonexistent, the perception on the part of the professorate is that this relationship exists (McCabe & Powell, 2004). In response to this perception, instructors do in fact lower their standards (Ryan, Anderson, & Birchler, 1980), although many will only admit that their colleagues engage in this behavior (McCabe & Powell; McSpirit, Kopacz, Jones, & Chapman, 2000). We theorize that instructors are willing to gradually, year by year, sacrifice academic standards for the incentive, real or imagined, of more satisfied students in the hope that such satisfaction will positively influence their salary, retention, and promotion, or, at the very least, lead to less emotional wear and tear caused by complaining students.

The incentive to give high grades to keep students satisfied may be reinforced by the influence of high school grade inflation. Between 1991 and 2003, high school GPAs rose by 0.20 to 0.26 (Woodruff & Ziomek, 2004) on a 4-point scale, a rate significantly higher than the rise in college GPAs over that same period (Rojstaczer & Healy, 2010). Given that many students now have never seen a C and rarely a B before entering

college, there is a tendency on the part of some students to equate B's with substandard performance and C's with failure. In a teaching environment where student satisfaction is of great importance, student perceptions as to what constitutes a good grade likely can influence a college instructor's grading habits.

CONCLUSIONS

Our data suggest that in the absence of oversight from leadership concerned about grade inflation, grades will almost always rise in an academic environment where professors sense that there are incentives to please students. In the Vietnam era, the incentive was created by external politics. Since the 1980s, the principal incentives likely have been a mix of the ability to enhance students' postgraduate prospects and, with the rise of the importance of student-based evaluations, faculty self-interest.

The gradual creep of grades wouldn't be so significant except for the fact that it has been so long-lived. When A is ordinary, college grades cross a significant threshold. Over a period of roughly 50 years, with a slight reversal from the mid-1970s to the mid-1980s, America's institutions of higher learning gradually created a fiction that excellence was common and that failure was virtually nonexistent.

The evolution of grading has made it difficult to distinguish between excellent and good performance. At the other end of the spectrum, some students who were once removed from school for substandard performance have, since the Vietnam era, been carried along. America's colleges and universities have likely been practicing some degree of social promotion for over 40 years. Evaluation has become so flawed that employers, graduate schools, and professional schools that try to use grades to identify outstanding prospects are likely often engaging in a futile exercise (Archer, Hutchings, & Johnson, 1998). Increasingly, they have to rely on standardized tests to make evaluations of student talent.

The implications of this transformation go well beyond being able to identify promising candidates for jobs and postgraduate schooling. When college students perceive that the average grade in a class will be an A, they do not try to excel (Babcock, 2010). It is likely that the decline in student study hours, student engagement, and literacy are partly the result of diminished academic expectations.

Significant exceptions to higher education's broken grading system come from schools that have established some oversight of grading practices. These efforts at self-regulation, although controversial in academia, are expected in most other professions. The standard practice of allowing professors free rein in grading has resulted in grades that bear little

relation to actual performance. It is likely that without the institution of grading oversight, either on a school-by-school basis or nationally, meaningful grades will not return to the American academy.

Acknowledgments

Jean Bahr, Steven Ingebritsen, and Nancy Weiss Malkiel provided helpful reviews of early versions of this article. Suggestions by one anonymous reviewer and by the editor, Lyn Corno, greatly improved the manuscript. Our work would not have been possible without the cooperation of many education journalists, university administrators, archivists and professors including Benny Amarlapudi, Kirk Baddley, Jennifer Ballard, Sheri Barrett, Laura Brown, Cheryl Browne, Edward Callahan, Neal Christopherson, Terri Day, Jennifer Dunseath, Cal Easterling, Jay Eckles, James Fergerson, Donna Gilleskie, Thomas Harkins, Sarah Hartwell, Earl Hawkey, Pamela Haws, Ann Henderson, Roy Ikenberry, Lauren Jorgensen, Brian Johnston, David Kane, Gary Kates, Isaac Krannick, Elizabeth Lieberman, Kristine Mascetti, Joseph Meyer, Kristen Noblit, Michael O'Donnell, David Oxtoby, Pat Parson, Ross Peacock, Kent Peterman, Jon Rivenburg, Alan Sack, Wayne Schneider, the late Marion Shepard, J. Kenneth Smail, Don Sprowl, Lawrence Summers, Rajah Tayeh, Daniel Teodorescu, and Richard Wagner.

References

- Adelman, C. (2004). *Principal indicators of student academic histories in postsecondary education, 1972–2000*. Jessup, MD: ED Pubs. (ERIC Document Reproduction Service No. ED483154)
- Archer, A. F., Hutchings, A. D., & Johnson, B. (1998). A case for stricter grading. *UMAP Journal*, 19, 299–313.
- Astin, A. W. (1971). *Predicting academic performance in college: Selectivity data for 2300 American colleges*. New York: Free Press.
- Aud, S., Hussar, W., Kena, G., Bianco, K., Frohlich, L., Kemp, J., et al. (2011). Table A-8-2: Actual and projected undergraduate enrollment in degree-granting 4- and 2-year post-secondary institutions, by sex, attendance status, and control of institution: Selected years, fall 1970–2020. In *The condition of education 2011: Indicator 8, undergraduate enrollment* (p. 172; NCES 2011-033). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office. Retrieved from http://nces.ed.gov/programs/coe/pdf/coe_hep.pdf
- Babcock, P. S. (2010). Real costs of nominal grade inflation? New evidence from student course evaluations. *Economic Inquiry*, 48, 983–996. doi:10.1111/j.1465-7295.2009.00245.x
- Babcock, P. S., & Marks, M. (2010). *The falling time cost of college: Evidence from half a century of time use data* (NBER Working Paper No. 15954). Cambridge, MA: National Bureau of Economic Research.
- Bar, T. R., Kadiyali, V., & Zussman, A. (2008). *Quest for knowledge and pursuit of grades: Information, course selection, and grade inflation at an Ivy League school* (Johnson School Research Paper Series No. 13-07). Ithaca, NY: Cornell University.
- Battersby, J. L. (1973). *Typical folly: Evaluating student performance in higher education*. Urbana, IL: National Council of Teachers of English. (ERIC Document Reproduction Service No. ED074519)
- Bayer, A. A. (1996). What is wrong with “customer”? *College Teaching*, 44, 82.
- Chace, W. M. (2009). The decline of the English department. *American Scholar*, 78, 32–42.

- College Board. (2007). *SAT data and reports, Table 2, mean SAT scores of college-bound seniors, 1967–2007*. Retrieved from http://www.collegeboard.com/prod_downloads/about/news_info/cbsenior/yr2007/tables/2.pdf
- DeMarzo, P. M., Fishman, M. J., & Hagerty, K. M. (2005). Self-regulation and government oversight. *Review of Economic Studies*, 72, 687–706.
- Dorans, N. (2002). *The recentring of SAT® scales and its effects on score distributions and score interpretations* (College Board Research Report No. 2002-11). New York: College Entrance Examination Board.
- Edson, F. G. (1951). Grade distribution in 80 Mid-western liberal arts colleges. *School and Society*, 73, 312–315.
- Ehrenberg, R. G. (2002). *Tuition rising: Why college costs so much*. Cambridge, MA: Harvard University Press.
- Ellis, L., Burke, D. M., Lomire, P., & McCormack, D. R. (2003). Student grades and average ratings of instructional quality: The need for adjustment. *Journal of Educational Research*, 97, 35–40.
- Faculty of Haverford College. (1972). Regular meeting, Education Policy Committee, Annex 1, April 20. Haverford College archives, Haverford, PA.
- Feldman, K. A. (1976). Grades and college students' evaluations of their courses and teachers. *Research in Higher Education*, 4, 69–111.
- Finkelmeier, T. (2010, January 27). Critics say grade inflation at UW-Madison lowers bar for students—and professors. *Madison Capital Times*. Retrieved from http://host.madison.com/ct/news/local/education/university/article_5adc496e-0ac6-11df-b737-001cc4c03286.html
- Frey, B. S. (2006). Giving and receiving awards. *Perspectives on Psychological Science*, 1, 377–388.
- Georgia Student Finance Commission. (2011). *Georgia's HOPE scholarship program overview*. Retrieved from March 17, 2011, from https://secure.gacollege411.org/Financial_Aid_Planning/HOPE_Program/Georgia_s_HOPE_Scholarship_Program_Overview.aspx
- Johnson, V. E. (2003). *Grade inflation: A crisis in college education*. New York: Springer.
- Juola, A. E. (1976, April). *Grade inflation in higher education: What can or should we do?* National Council on Measurement in Education annual meeting, San Francisco, CA. (ERIC Document Reproduction Service No. ED129917)
- Juola, A. E. (1980). *Grade inflation in higher education-1979. Is it over?* East Lansing: Learning and Evaluation Service, Michigan State University. (ERIC Document Reproduction Service No. ED189129)
- Kaplan Test Prep and Admissions. (2010, March 2). *Course selection in the rate-my-professors era* (Press release). Retrieved from <http://press.kaptest.com/category/press-releases/company-news>
- Kezim, B., Pariseau, S. E., & Quinn, F. (2005). Is grade inflation related to faculty status? *Journal of Education for Business*, 80, 358–363.
- King, J. E. (2000). *Gender equity in higher education: Are male students at a disadvantage?* Washington, DC: American Council of Education.
- King, J. E. (2010). *Gender equity in higher education: 2010*. Washington, DC: American Council of Education.
- Kuh, G. D., Kinzie, J., Schuh, J. H., & Whitt, E. J. (2005). *Student success in college: Creating conditions that matter*. San Francisco: Jossey-Bass.
- Kuh, G. D., & Hu, S. (1999). Unraveling the complexity of the increase in college grades from the mid-1980s to the mid-1990s. *Educational Evaluation and Policy Analysis*, 21, 297–320.

- Kutner, M., Greenberg, E., & Baer, J. (2006). *A first look at the literacy of America's Adults in the 21st century* (NCES 2006-470). Jessup, MD: U.S. Department of Education. (ERIC Document Reproduction Service No. ED489066)
- Levine, A., & Cureton, J. S. (1998). *When hope and fear collide: A portrait of today's college student*. San Francisco: Jossey-Bass.
- Mansfield, H. C. (2001, April 6). Grade inflation: It's time to face the facts. *Chronicle of Higher Education*, p. B24.
- Marsh, H. W., & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist*, 52, 1187-1197.
- McCabe, J., & Powell, B. (2004). "In my class? No." Professors' accounts of grade inflation. In W. E. Becker & M. L. Andrews (Eds.), *The scholarship of teaching and learning in higher education* (pp. 193-220). Bloomington: University of Indiana Press.
- McSpirit, S. J., & Jones, K. E. (1999). Grade inflation rates among different ability students, controlling for other factors. *Education Policy Analysis Archives*, 7.
- McSpirit, S. J., Kopacz, P., Jones, K. E., & Chapman, A. D. (2000). Faculty opinion on grade inflation: Contradictions about its cause. *College and University*, 75, 19-25.
- Meyer, M. (1908). The grading of students. *Science*, 28, 243-250.
- National Center for Education Statistics. (2005a). *Youth indicators 2005: Trends in the well-being. Figure 30. Percentage of 16- to 24-year-old college students who were employed, by attendance status and hours worked per week: October 1970 to October 2003*. Retrieved from <http://nces.ed.gov/programs/youthindicators/Indicators.asp?PubPageNumber=30>
- National Center for Education Statistics. (2005b.) *Digest of education statistics, list of tables and figures 2005*. Retrieved from http://nces.ed.gov/programs/digest/2005menu_tables.asp
- National Center for Education Statistics. (2008a.) *Table 216. Total fall enrollment in degree-granting institutions, by level of enrollment and state or jurisdiction: 2004, 2005, and 2006*. Retrieved from http://nces.ed.gov/programs/digest/d08/tables/dt08_216.asp
- National Center for Education Statistics. (2008b.) *Table 204. Enrollment rates of 18- to 24-year-olds in degree-granting institutions, by type of institution and sex and race/ethnicity of student: 1967 through 2007*. Retrieved from http://nces.ed.gov/programs/digest/d08/tables/dt08_204.asp
- National Center for Education Statistics. (2009a). *2006 SAT data for all institutions downloaded from the IPEDS data center*. Retrieved from <http://nces.ed.gov/ipeds/datacenter>
- National Center for Education Statistics. (2009b). *Table 329. Number of applications, admissions, and enrollees; Their distribution across institutions accepting various percentages of applications; and SAT and ACT scores of applicants, by type and control of institution: 2008-09*. Retrieved from http://nces.ed.gov/programs/digest/d09/tables/dt09_329.asp
- National Center for Education Statistics. (2009c). *Table 198. Total first-time freshmen fall enrollment in degree-granting institutions, by attendance status, sex of student, and type and control of institution: 1955 through 2008*. Retrieved from http://nces.ed.gov/programs/digest/d09/tables/dt09_198.asp
- Perrin, N. (1998, October 9). How students at Dartmouth came to deserve better grades. *Chronicle of Higher Education*, p. A68.
- Perry, W. M. (1943). Are grades and grading systems comparable from institution to institution? *Registrar's Journal*, 18, 159-165.
- Rojstaczer, S., & Healy, C. (2010). Grading in American colleges and universities. *Teachers College Record*, ID Number 15928.
- Ryan, J. J., Anderson, J. A., & Birchler, A. B. (1980). Student evaluations: The faculty responds. *Research in Higher Education*, 12, 317-333.

- Saenz, V. B., & Barrera, D. S. (2007). *Findings from the 2005 College Student Survey (CSS): National Aggregates*. Los Angeles: UCLA Higher Education Research Institute.
- Seldin, P. (1984). *Changing practices in faculty evaluation*. San Francisco: Jossey-Bass.
- Seldin, P. (1993, June 12). The use and abuse of student ratings of professors. *Chronicle of Higher Education*, p. A40.
- Spiegel, M. R., & Stephens, L. J. (1998). *Schaum's outline of statistics* (3rd ed.). New York: McGraw-Hill.
- Stassen, M. (2001). Non-returning first-year students: Why they leave and where they go. *UMass Assessment Bulletin*, 4(2), 1–4.
- Suslow, S. (1976). *A report on an interinstitutional survey of undergraduate scholastic grading 1960s to 1970s*. Berkeley: Office of Institutional Research, University of California. (ERIC Document Reproduction Service No. ED129187)
- Vars, F. E., & Bowen, W. G. (1998). *Scholastic Aptitude Test scores, race, and academic performance in selective colleges and universities*. In C. Jencks & M Phillips (Eds.), *The Black-White test score gap* (pp. 457–479). Washington, DC: Brookings Institution Press.
- Weinberg, S. L. (2007). Grade inflation: An examination at the institutional level. In S. S. Sawilowsky (Ed.), *Real data analysis* (pp. 315–323). Charlotte, NC: Information Age.
- Weinberg, B. A., Fleisher, B. M., & Hashimoto, M. (2007). *Evaluating methods for evaluating instruction: The case of higher education* (National Bureau of Economic Research Working Paper No. 12844). Cambridge, MA: National Bureau of Economic Research.
- Woodruff, D. J., & Ziomek, R. L. (2004). *High school grade inflation from 1991 To 2003* (ACT Research Report Series, 2004-4). Iowa City, IA: ACT.
- Zirkel, P. A. (1994). Students as consumers. *Phi Delta Kappan*, 76, 168–171.

APPENDIX

Grade distribution data came from the following schools, in addition to data from 33 and 11 anonymous schools from Perry (1943) and Suslow (1976), respectively.

Abraham Baldwin, Alaska–Anchorage, Alaska–Fairbanks, Appalachian State, Auburn, Augusta State, Baylor, Belmont, Benedict, Benedictine, Bowdoin, Brown, Bucknell, Cabrini, California–Berkeley, Cal Poly–San Luis Obispo, Carleton, Catholic University of America, Central Florida, Central Michigan, Centre, Charleston, Citadel, Clarion, Clemson, Coastal Carolina, College of New Jersey, Colorado–Boulder, Colorado College, Colorado–Colorado Springs, Colorado–Denver, Columbus State, Connecticut, Cornell, CSU–Fresno, CSU–LA, CSU–Long Beach, CSU–Northridge, CSU–San Bernardino, Dalton State, Dartmouth, Delaware, DePauw, Doane, Dominican, Duke, Elon, Emory, Fairmont State, Ferris State, Fisk, Florida–Gainesville, Florida International, Florida State, Framingham State, Francis Marion, Furman, George Mason, Georgetown, Georgia, Georgia State, Georgia Tech, Grinnell, Harvard, Hawaii–Hilo, Hawaii–Manoa, Hawaii–West Oahu, Hope, Idaho State, Illinois–Champaign Urbana, Illinois–Chicago, Illinois Wesleyan, Indiana–Bloomington, Indiana State, Indiana Purdue–Fort Wayne, Indiana Wesleyan, Ithaca, James Madison, Kennesaw State, Kent State, Kenyon, Knox, Lander, Linfield, Louisiana State–Baton Rouge, Macalester, Messiah, Methodist,

Michigan, Michigan Tech, Miami (Ohio), Minnesota–Twin Cities, Mississippi University for Women, Missouri–Columbia, Missouri State, Missouri Western, MIT, Montana–Missoula, Montclair State, Nazareth, Nebraska–Lincoln, Nevada–Las Vegas, Nevada–Reno, New Orleans, North Alabama, North Carolina–Asheville, North Carolina–Chapel Hill, North Carolina–Wilmington, North Dakota State, Northern Arizona, Northwestern, Oakland, Oberlin, Ohio, Oral Roberts, Oregon Institute of Technology, Pennsylvania State, Pennsylvania, Pomona, Portland State, Princeton, Purdue, Reed, Rhodes, Rice, Roanoke, Rowan, Rutgers, Salisbury, San Jose State, Skidmore, Smith, Sonoma State, South Carolina–Columbia, South Florida–Tampa, Southeastern Louisiana, Southern Mississippi, Southern Polytechnic, Southern Utah, Spelman, St. Michaels, Stanford, SUNY–Purchase, Tarleton State, Tennessee–Chattanooga, Tennessee Tech, Texas A&M, Texas–Arlington, Texas State, Towson, Utah–Salt Lake City, Utah Valley, Virginia Commonwealth, Washington State, West Georgia, Western Carolina, Western Kentucky, Western New England, Wheeling Jesuit, Whitman, William and Mary, William Paterson, Williams, Winona State, Winthrop, Wisconsin–Madison, and Yale.

STUART ROJSTACZER is a retired professor of geology, civil engineering, and environment at Duke University. He works as a consultant on water issues and has research interests in the evolution of pedagogy and grading in colleges, the evolution of permeability in the continental crust, and the human use of water at local and global scales. Recent publications include: Rojstaczer, S., Ingebritsen, S. E., & Hayba, D. O. (2008). Permeability of continental crust influenced by internal and external forcing, *Geofluids*, 8, 1–12; and Rojstaczer, S., & Healy, C. (2010). Grading in American colleges and universities. *Teachers College Record*, ID Number 15928.

CHRISTOPHER HEALY is an associate professor of computer science at Furman University. His research interests include timing analysis for real-time and embedded systems, programming languages, computer science education, and grading issues and pedagogy in higher education. Recent publications include: Healy, C. (2010). International computing issues as a freshman seminar. *Computing Sciences in Colleges*, 25, 143–148; and Mohan, S., Mueller, F., Hawkins, W., Root, M., Healy, C., & Whalley, D. (2010). Parametric timing analysis and its applications to DVS. *ACM Transactions on Embedded Computing Systems*, ID Number 10.1145/1880050.1880061