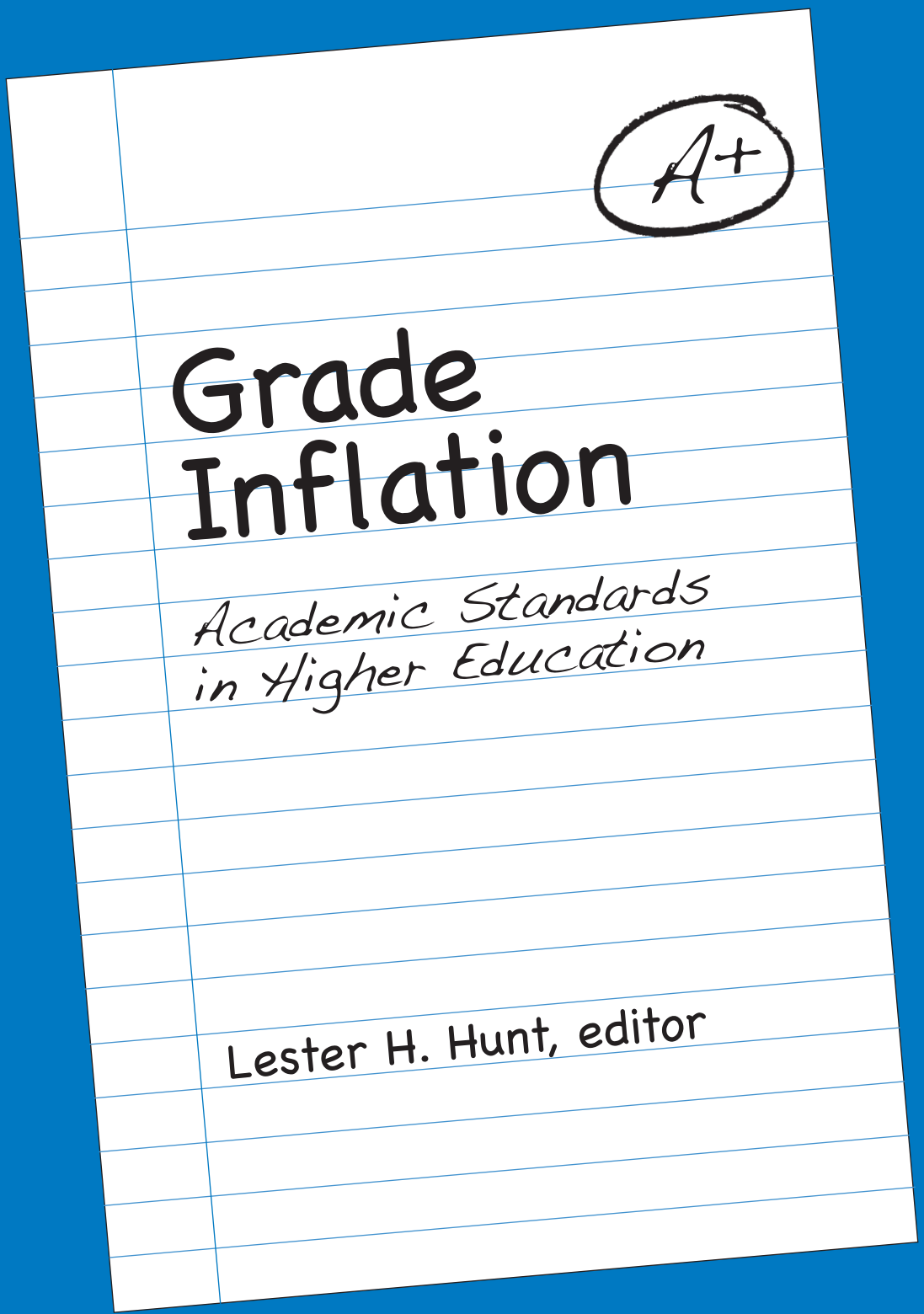

A+
Grade Inflation
Academic Standards in Higher Education
Lester H. Hunt, editor

Grade Inflation

This page intentionally left blank.

Grade Inflation

Academic Standards in Higher Education

EDITED BY

Lester H. Hunt

State University of New York Press

Published by State University of New York Press, Albany

Printed in the United States of America.

For information, contact State University of New York Press, Albany, NY.
www.sunypress.edu

Production by Diane Ganeles
Marketing by Michael Campochiaro

Library of Congress Cataloging-in-Publication Data

Grade inflation : academic standards in higher education / edited by Lester H. Hunt.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-7914-7497-6 (hardcover : alk. paper)
1. Grading and marketing (Students)—United States.. 2. Education, Higher—Standards—United States. I. Hunt, Lester H., 1946–

LB2368.G73 2008
371.27'2—dc22

2007034088

CONTENTS

FOREWORD

Grade inflation is one of those topics that initially seems clear and simple but becomes murkier and more confusing the longer you think about it. Consider the following hypothetical scenario: At Great State University (GSU) the same faculty teach the same courses, assign the same homework, administer the same examinations, and use the same grade standards and definitions year after year. There are no fraternity files or other homework or exam files to be shared with incoming students. Each year the new class of GSU freshmen exhibits the same distribution of high school grades and test scores as prior entering classes. For years this has produced a symmetrical distribution of GSU grades (let's say, equal percentages of As and Fs; equal percentages of Bs and Ds; and more Cs than any other single grade). So for years the mean, median, and modal grade has been C, or 2.0. That comports well with "common sense" and common understandings: Some students are outstanding, some fail to understand or perform at all, and everyone else is somewhere in between. A "C" grade is widely accepted to mean "average," and the average student should, almost by definition, be expected to get an average grade.

Suppose, in some future year, it is noticed that everything else has remained the same, but the GSU average grades have been rising for several years in a row—fewer low grades and more high grades, shifting the averages upward. Is this, by itself, evidence of what most people mean when they talk about "grade inflation" and "relaxed standards"? Probably so. On the other hand, this hypothetical situation has never prevailed anywhere. All the things I postulated to be unchanging at GSU are changing constantly for real universities, and it is remarkably

difficult to tease out and identify unambiguously a "grade inflation" that would be universally recognized as being unjustified or inappropriate. Older faculty retire; new, better-trained faculty are hired (yes, better—teacher training, which almost none of us ever received, is now nearly universally available in top graduate schools); college courses and curricula change (for the better, one hopes); universities try consciously to improve teaching and learning outcomes; universities train K–12 teachers and influence high school curricula with a goal of improving teaching and learning at that level; and admissions officers try consciously to select and admit only the best-prepared applicants. All of these changes can and do affect academic performance in college and, therefore, might plausibly be expected to affect average college grades. Or should they?

Consider another scenario: GSU finds (for whatever reason) that the overwhelming majority of its applications for freshman admission are from students who are in the top 5 percent of their high school class and who have test scores placing them in the top 5 percent of all test takers nationally. When the best of these exceptionally bright, well-prepared students are admitted, what should the faculty expect? Should they expect better performance in GSU classes? If so, how should they respond? Should they give higher grades reflecting that performance, or should they ratchet up the difficulty of their courses as far as necessary to guarantee the historic percentage of Ds and Fs? The latter is always possible to do. But, as a matter of fairness, not to mention public policy, is it reasonable for GSU to take in the best of our high school graduates and label some fraction of this elite group as "failures" simply because they were not "the best of the best"?

This goes to the core question: What is the goal or purpose of grading? Are grades given to discriminate among the students, always striving to produce some symmetrical (perhaps a normal) distribution of grades so graduate schools and employers can identify the "best" and "worst" of the graduates, no matter how good the "worst" may be? This problem is almost universally addressed in American graduate schools by compressing the grade scale: At most graduate schools, outstanding performance and corresponding "A" grades are expected. Indeed, an average "B" grade is required for satisfactory progress and continuation as a graduate student. This is explicitly because graduate schools accept only the highest-performing bachelors graduates ("All our graduate students are above average!"), and anything lower

than a "C" grade is simply punitive because that amounts to a failing grade. So, in this sense, "grade inflation" is not considered a defect of graduate education, it is a design feature! Why should grading for GSU undergraduates be any different if all of GSU's admissions are from the highest-performing slice of the high school graduates? So, reasoning by analogy with the de facto grading practices in graduate schools, should undergraduate grades simply certify that the students have either mastered or not mastered a certain body of knowledge and either exhibit or do not exhibit a defined set of skills or capabilities? If so, that amounts to a "pass-fail" system, in which case only two grades (A and F or, indeed, A and B) would suffice. Is it possible that the pass-fail system is appropriate in some disciplines and not in others? And so it goes. I could pose more scenarios and questions illustrating the complexity of this topic, but, after all, that is the subject of this book, and it is better addressed by the expert authors than by me.

Still, when all the arguments and rationalizations are over, a core issue must be addressed: How can we make certain, and how can we assure ourselves, the students, and the public at large that grades truly do represent something meaningful, whatever that may be—that we have "standards" on which people can rely? If there are elements of the "commonsense" understanding of grade inflation in our grading practices, then shouldn't we identify, address, and either eliminate or explain them? This, I believe, is the central, motivating concern.

To that concern, I would like to add another related question that places apparent grade inflation in direct conflict with a major public policy issue. Colleges and universities are under great pressure, as a matter of public accountability, to increase their graduation rates. When most of the authors and readers of this volume were in school, college graduation rates were much lower than they are today. I frequently hear alumni relate how, in one of their first classes as a new freshman, the professor told them, "Look to your right, and look to your left: Only one of you is going to graduate." This admonition was undoubtedly intended to be motivational: "Study hard because failure is a real possibility." And the data bear this out. In the 1960s, graduation rates at Wisconsin and other large public universities were less than 50 percent—often far less. But with rising costs for K–12 and postsecondary education, failure is no longer considered acceptable by the public. Legislatures and trustees are increasingly holding us accountable for improving our graduation rates, and we are doing

it. Today, the median graduation rate of the public AAU (American Association of Universities) universities is 73 percent, and the median graduation rate for the private AAU universities is 88 percent. If we admit only those students our admissions officers believe can succeed in college, if we take it as our responsibility to help those students succeed (almost the definition of teaching!), and if we monitor their progress by requiring them to maintain a *minimum* C average (grade point average [GPA] 2.0) to continue in school and to graduate, then the average GPA of the student body will necessarily be *greater than* 2.0. That is a mathematical certainty. Paradoxically, it means that the average student must necessarily maintain above-average grades. So some amount of "grade inflation" is a necessary consequence of doing our jobs well, and "average" is not a term properly applied to the grade that defines the bottom of the acceptable range for graduation. When the graduation rate rises above 50 percent, the average student is, in this sense, automatically above average!

University faculty have a dual responsibility: On the one hand, they are expected to define and maintain high standards of academic excellence. The grades they assign and the degrees they award are supposed to certify to the public that those standards were met. At the same time, faculty are being held accountable for raising graduation rates: "We're paying you to produce degrees, not Fs!" Clearly these expectations are in conflict, at least to the extent that we allow our grading practices simply to evolve without periodic examination and public discussion.

When I was provost, I asked our Academic Planning Council to undertake an examination of all aspects of our grading practices for exactly these reasons of self-examination and public assurance. Our Office of Institutional Research examined several years of grade data and, through a very sophisticated analysis, concluded that our average GPA went up very slightly over the period 1990–1998, and that only about a third of the increase was attributable to the increasing academic preparation of our students. For example, students in 1998 received grades averaging about 0.2 GPA points higher than similarly prepared students received in 1990. Only about 0.06 GPA points of this increase could be "explained" by detailed subtleties of student demographics, leaving 0.14 GPA points of the increase to be explained by other factors. This finding was based on averages over all schools and colleges and over all undergraduate levels. Great variation was found

at the individual school and college level, and for different slices of the distribution of student preparation, so no single conclusion could be reached for the entire university. Neither could we reach agreement on the cause or meaning of the increase for any individual school, college, or department.

That does not mean the exercise was a failure, however. In the process of searching for evidence of grade inflation, we discovered some other things about grades and grading that were of concern and would not have been addressed without the study. One is the grade variations among sections in multisection courses. Two students with, for all practical purposes, identical performance but in different sections of the same course should not get dramatically different grades. And that was happening in some courses. All other things aside, students should not have to depend on the luck of the draw as to which section they get in. The faculty agreed with this proposition and implemented measures to address the problem. We also examined some mathematical artifacts inherent in our seven-level quantized grading system (A, AB, B, BC, C, D, F) and addressed those in some of our rules that contain numerical GPA thresholds and cutoffs. In the end, even though we did not definitively identify or "solve" any problems of "grade inflation" per se, we did gain a much healthier understanding of our grading policies and practices, and we put ourselves in a much better position to be accountable for them. I heartily recommend that all universities periodically undertake studies of this sort to assure themselves and the public that they are continuing to maintain appropriate and fair standards while simultaneously educating and graduating more students.

Finally, I would like to commend Professor Lester Hunt and the Wisconsin Association of Scholars for organizing this important conference, and to thank all the contributors and participants. We need to engage in this sort of public self-examination on many topics, and "grade inflation" is an excellent starting point.

John D. Wiley, Chancellor
University of Wisconsin-Madison

NOTE

This foreword is based on Chancellor Wiley's welcoming remarks at the opening of the Conference on Grade Inflation and Academic Standards at the University of Wisconsin-Madison on October 11, 2003.

This page intentionally left blank.

PREFACE: WHAT ARE THE ISSUES?

The inspiration for this book originated in a nationwide discussion that was ignited, during 2002 and 2003, by media reports of grading practices at elite eastern universities, including Harvard and Princeton. The practices, as reported, involved awarding what most Americans would intuitively regard as excessive numbers of As. Among the sparks that set the controversy ablaze were comments by Harvard political theorist and "grade inflation" critic Harvey Mansfield.[1] More fuel was added to the discussion when education theorist Alfie Kohn published an article, a lightly revised version of which is printed here, in the November 8, 2002, issue of *The Chronicle of Higher Education*. In it, he asserted that, contrary to what much of the recent media comment suggested, there may actually be no such thing as grade inflation in American higher education today.

With a grant from the Wisconsin Association of Scholars, I organized a conference on "grade inflation" and closely related issues, which was held at the University of Wisconsin-Madison on October 11, 2003. Almost all the chapters in this book are derived from (and in some cases are radically revised versions of) papers presented at that conference. The purpose of this book, as of the conference from which it arose, is to examine the issues surrounding the idea of grade inflation from as many different perspectives and methodologies as possible. Some of them are strongly empirical, and some are entirely conceptual. Some are scientific, while others are openly polemical or moralistic. Some maintain that grade inflation is an epidemic problem in need of a radical solution, while others doubt that it exists. A reader who dives right into these eddying currents is apt to fall into a state of confusion.

What I would like to do here is minimize this confusion if I can. I will try to say, as briefly and clearly as I can, what the basic issues are that our authors are treating and how the different essays are logically related to one another. Who is disagreeing with whom, and about what? Which of the views presented conflict, and which are complementary to one another?

First, I should say what *sorts* of issues are involved here. Purely for convenience—without, I hope, committing myself to any deep philosophical theories—I will divide the controversial issues into three broad categories: *conceptual* issues, *normative* issues, and *empirical* issues. Conceptual issues include questions of the proper analysis or definition of concepts, which for present purposes can be thought of as the meanings of words. For instance, the question of what *war* is, insofar as it is a question of the proper definition of the word "war," is a conceptual issue. So is, with the same qualification, the question of whether a revolution is a sort of war, and of whether America's "war" on terror is a war, or its "war" on drugs. Although it is always possible to frame conceptual issues as "merely verbal" ones, they are sometimes very important. The concepts (or words) people use to deal with the world tend to bend and shape their thoughts and actions. Once we start to think of a campaign against drugs as a war, we are willing to use methods that would not otherwise be permissible, or perhaps even thinkable. We can do things, and with a clear conscience, that we would not otherwise be willing to do at all. One sort of conceptual issue that can become important is whether a concept is expressed by a metaphor (either live or dead) that interacts with the concept in confusing ways.

Normative issues include questions of what people should or should not do, or of what would be good or bad things to do. Obviously these issues go to the very heart of the controversy about grade inflation. The reason there is disagreement is that some people think that professors are doing something they should not be doing, or that there is some other way of conducting their affairs that would be better.

Finally, empirical issues are questions about what (regardless of what should or should not be the case) is observably going on in the world. One legitimate input into these issues is what might be called "armchair empiricism," or reliance on what one has already observed. Most of the people who read this volume assign grades as part of their professional activities, and they know by direct observation what they themselves do, and what some of their colleagues do. Obviously

to make much headway with these issues we need more and better empirical evidence, including primary historical documents, questionnaire responses, grade transcripts, and perhaps controlled experiments.

In addition, I will comment on two "bottom-line" issues, issues that rest on one's solution to several of the more fundamental ones.

CONCEPTUAL ISSUES

How should we define "grade inflation"? Alfie Kohn defines grade inflation as " an upward shift in students' grade point averages without a similar rise in achievement." Closely parallel definitions are proposed by Clifford Adelman and Harry Brighouse. As Kohn suggests, this definition would seem to imply that increases in grades do not constitute inflation if they happen because students are turning in better assignments, or because there are fewer requirements that force them to take courses outside of their areas of strength, or because it is easier to drop courses in which they are not doing well. If students are doing better in the courses for which they received grades, he seems to be saying, then perhaps grades ought to rise. The idea is that grade "inflation," if the term is to be meaningful, must be closely analogous to economic inflation. Adelman, for instance, comments that grade inflation, if it exists, would consist in teachers *paying* a higher and higher *price* for the *same product* from students.

Of course, as is generally the case with hotly contested ideas, other definitions are possible. Richard Kamber, in "Understanding Grade Inflation," defines grade inflation as the "reduction in the capacity of grades to provide reliable and useful information about student performance as a result of upward shifts of grading." Note that Kamber's proposed definition involves an assumption to the effect that the concept of grade inflation, if it makes sense, is the concept of a malfunction, and that it cannot be neutral about what this malfunction is. On his analysis, the malfunction is a failure in communicating information to students. Other definitions might imply that the relevant malfunction would be that some grades are not justified because the quality of the students' work does not deserve those particular grades. One immediate implication of Kamber's definition is that an upward shift in grades can be inflationary and still cause problems (contrary to what Kohn states and Adelman and Brighouse imply) even if it is accompanied by an increase in the quality of the asssignments that students

turn in. The reason is that the evil of grade inflation, in Kamber's analysis, lies in something that it has in common with grade *de*flation: namely, that it is an instance of grade *con*flation. I will return to this idea momentarily.

Brighouse offers arguments that in effect are rejoinders to this line of reasoning. He maintains that students are not misled by inflated grades because, in the event that they find themselves in a class with inflated grades, they "discount" them accordingly. Note that this is just what economic agents do regarding price inflation. If bankers who make loans know how much the money supply will be inflated, they require interest rates high enough to offset the diminished value of the dollars in which their loans will be repaid. In addition, Brighouse claims that that the conflation of the highest levels of achievement into one grade affects relatively few students (namely, the very best), and that these tend to be the ones who are least interested in grades.

Is "grade inflation" a useful term? On the basis of his strongly economic analysis of the concept of grade inflation, Adelman concludes that it would not apply meaningfully to higher education. After all, there is no *same product* for which teachers could be paying either the same price or different prices. The papers and exams they were evaluating in 2002 were quite different from those that were being evaluated in 1972, because knowledge has grown since then. Interestingly, Kamber reaches a similar conclusion—that is, that the metaphor of "inflation" is not very helpful in understanding grading trends—though he does so on entirely different grounds. In his analysis, the difference in quality between student performance in 1972 and today is irrelevant to what grades we should be giving now. For him, the crucial difference between economic inflation and grade "inflation" is that the problems they cause are entirely different. Price inflation is an evil (if and when it is) because it causes an inefficient allocation of resources and an unjust redistribution of wealth and income, and none of these bad effects will happen if inflation is uniform and universally anticipated. Grade inflation is bad, according to Kamber, because of the way in which grading is *different* from pricing. Grades, unlike prices, constitute a system that is sealed at both ends, so the general upward or downward shifts result in phenomena being bunched up at one end of the scale: A refers to what used to be A work, but also to what used to be B work, so that it becomes impossible to distinguish between

those two levels of achivement any longer. This effect occurs, he maintains, even if the upward shift is uniform and anticipated.

NORMATIVE ISSUES

What is the proper purpose of grading? As we have seen, our answers to the question of what grade inflation is might rest on answers to a more fundamental question, of what grading is supposed to accomplish. Should grading be used to distinguish between the quality of performance of the students in a given class, program, or school? Or should it be used to assign absolute standards of value? Or should it inform prospective employers or admissions committees as to how well students have done? Or should it be used primarily to motivate students? Alfie Kohn speaks against the first of these answers, that grading should sort students out into levels of achievement, on the grounds that it creates a conflict of interest among students, a race to be at the head of the class. John D. Wiley points out that the second sort of policy—assigning absolute levels of achievement—could explain graduate schools in which the average grade is an A (if you are not outstanding, then you should not be in graduate school). Kohn criticizes the last of these functions, that of providing motivation, on the basis of his theory that such "extrinsic" motivation crowds out the "intrinsic" motivation (learning for its own sake) which, in his view, learning requires.

Kamber's analysis, as we have seen, assumes that none of the previously mentioned purposes of grading is the fundamental one. The basic purpose of grading, in his view, is to communicate information to students about how well they are doing. Brighouse agrees that this is at least *one* important purpose of grading. He points out that grades are important cues that students can use as they allocate study time to different assignments and courses. While Kamber argues that inflation is a problem mainly because of the informing function of grades, Brighouse argues on the contrary that it is most likely to be problematic in relation to a completely different function, namely, that of motivating students. If As become routine, then they lose some of their power to inspire effort, while if Cs become rare, then they are more likely (when they are given) to discourage students and crush their desire to try harder.

An additional potential purpose of grading is discussed by Francis Schrag. It might be called "socialization." Grading is done by applying

norms of conduct to human behavior—this is good, that is not good. One thing that grading might do is convey these norms to others, to imprint these norms on the minds of the students.

What ethical standards are relevant to grading practices? The two authors that focus on this issue are Schrag and (in "Grade 'Inflation' and the Professionalism of the Professoriate") Mary Biggs. They give sharply contrasting answers. Schrag explores the implications of certain egalitarian conceptions of social justice and attempts to construct some grading practices that are compatible with his preferred variety of egalitarianism. The basic problem he is trying to confront is that grading, by its very nature, seems to treat people unequally. Biggs, on the other hand, approaches the subject from the point of view of professional ethics, in particular a version of professorial professional ethics that applies to the professoriate. In her view, the fundamental obligation of this ethic is to promote excellence. Obviously her fundamental principle, excellence, is very different from Schrag's, which is equality, and it promises to lead those for whom it is a fundamental principle in quite a different direction.

EMPIRICAL ISSUES

Are there any demonstrable nationwide trends in grading practices? If there are, what are they? Obviously one of the issues at stake in the debate about grade inflation is that of what has actually happened to grades over the years. Kamber ("Understanding Grade Inflation") discusses evidence that the first wave of grade inflation in the United States was not in the 1960s, as often assumed, but in the pre-Civil War period, when new colleges proliferated and fledgling institutions could ill afford to fail students. The later nineteenth century, he believes, saw a gradual tightening of standards. There also is evidence, he notes, that there was an inflationary surge, nationwide, in the late 1960s and early 1970s. Citing studies published in 1976 and 1979 by Avro Juola, he finds a marked and strikingly universal upswing in grades from 1968 to 1974, with a slight decline in grades after the peak year 1974.

Clifford Adelman discusses the results of three nationwide longitudinal studies conducted by the National Center for Education Statistics of the U.S. Department of Education, which followed members of high school graduating classes of 1972, 1982, and 1992 for a minimum of twelve years, using grade transcripts from over 2,500 educational

institutions. He shows that these data show only rather subtle and complex changes from decade to decade, and the changes do not necessarily go in the direction that one might have expected. The high school graduating class of 1982 got slightly *fewer* As and slightly *more* Cs than the high school graduating class of 1972. Overall, the best data do not show a single-direction, nationwide trend throughout the three decades following 1972.

One feature of these two accounts, that of Kamber and that of Adelman, that is very salient, at least to me, is that each is perfectly consistent with the other. And yet, oddly enough, they are rhetorically positioned in opposition to each other. Kamber thinks grade inflation is a real problem, while Adelman is skeptical about whether it exists at all. How is this possible? Part of the answer, I think, lies in the conceptual differences we have already seen. If we ask whether economic inflation is (note the present tense) a problem, then we are asking about current shifts. What happened during the era of the Vietnam War is ancient history, irrelevant to the matter in question. On Kamber's view, that is not true of grade inflation—if grades were conflated during the Vietnam period and never came close to returning to earlier configurations, that can be a problem now.

What sorts of causal factors result in grades (sometimes) shifting upward? The answer one gives to this question will affect whether one thinks a given grading trend is inflationary, and it also will affect one's views on other relevant issues as well, such as whether reform is feasible and desirable, and which reforms one should prefer.

Kohn and Wiley point out that grades might go up because teachers are doing their jobs more effectively. As Wiley points out, helping students succeed is almost the definition of teaching. As we have seen, some of our authors are inclined to deny that grades that change for this reason are inflationary.

Kamber points out that upward shifts also can be the result of ideology. Teachers may want to show solidarity with students. Schrag indicates, in a somewhat indirect way, another way in which grading can be infuenced by ideology. He describes several grading policies that accord with "luck egalitarianism"—roughly, the idea that what you get in life should not be a function of luck (as it would be if it were a result of your genetically determined abilities). All of the practices he describes involve allowing students to do extra-credit work to raise their grades. Obviously this would, other things being equal,

tend to raise the average grade. Further, as Schrag points out, they also would tend to muddy the informational content of grades in just the way that critics like Kamber identify with grade inflation, making it harder to tell from grades alone how good the students' actual performance is.

Kamber and Wiley speculate that upward shifts also can be influenced by pressures to increase graduation rates. If we lower our standards then, by definition, more students are doing satisfactory work. Kamber, Brighouse, Biggs, and Hunt also argue that the advent of student evaluation of teaching in the late 1960s probably exerted an upward pressure on grades. Another possible cause discussed by Kamber and Biggs is enhanced time constraints on college teachers, such as an increased demand that they produce scholarly work of their own. For various reasons, lax grading consumes less time than rigorous grading. If you give a D or an F, and sometimes a C, then you will probably have to justify it in detail, while no one is going to demand that you justify giving an A. Professors who can ill afford to spend time meeting with angry students and parents have a reason to give plenty of As.

One possible causal factor that is denied by Biggs and Kamber is the frequently heard allegation that admission of inadequately prepared nonwhite students has led intimidated (white liberal) teachers to go easy on inadequate work. The modern wave of grade inflation began in the 1960s, when minority students were only 8 percent of the student population, and studies show them getting lower grades than white students with the same Scholastic Aptitude Test (SAT) scores.

Mary Biggs, in her two contributions to this book, traces grade inflation (or, as she prefers to say, grade "conflation") to two different causes. In "Grade 'Inflation' and the Professionalism of the Professoriate," she says that the cause of grade inflation is the faculty. This might sound like a mere witticism to some, but given that her approach (at least in this chapter) is fundamentally ethical, it is just what she (to be consistent) ought to say. As Nietzsche said long ago, morality is a theory of causes. If you are going to blame someone for some bad thing that has happened, then you must assume that the reason the bad thing happened was that person's act of will (including the person's abdications of the responsibility to choose), and you must suppose that the reason the thing happened is *not* to be found in the many things that the person cited as excuses for what was done. On Biggs's view,

the real reason grade inflation actually happens is that professors put personal profit above professional ethics.

Interestingly, in "Fissures in the Foundation: Why Grade Conflation Could Happen," Biggs gives a rather different sort of explanation of grade inflation. There she argues that the ultimate cause of grade inflation is literally profound, namely, the decay of the intellectual *foundations* of grading. The sorts of causes that are usually mentioned (such as student evaluation of teaching) are at best superficial and intermediary causes. Grading as we know it today developed in tandem with a sort of scientism that held that the mind can be measured with scientific precision. As these ideas (deservedly) fell out of favor, and no new orthodoxy moved in to replace it, instructors lost confidence in their ability (or right) to assign grades. Insofar as Biggs is offering a historical explanation of grade conflation in the "Fissures" chapter, an unsympathetic reader might jump to the conclusion that her position here is incompatible with the moralizing sort of explanation she offers in the "Professionalism" chapter. If the cause lies in the historical situation, then the faculty cannot help what they do and cannot be blamed. But there are fairly straightforward ways in which Biggs can make the two explanations mutually consistent. For instance, she might say that the fissures in the intellectual foundations of grading simply removed a certain cultural support that facilitates conscientious grading. As the foundations crumble, something that once was a routine undertaking begins to require individual virtue and so becomes a test of one's character. If one fails the test, one can be blamed for doing so. Biggs's historical explanation is offered to show why grade inflation *can* happen, and her moralizing explanation is offered to show why it *does* happen.

Kamber, we might note, expresses skepticism about the historical explanation that Biggs offers in the "Fissures" essay. He argues (in "Understanding Grade Inflation") that most of the professors who assign grades were untouched both by the intellectual principles of the psychometrics movement and those of its opponents.

THE BOTTOM LINE

Finally, there are the two bottom-line issues. They are fundamentally normative in nature and rest on many of the issues described earlier, including the other normative ones. They are: Is grade inflation a problem? And, if it is, is there anything that can be done about it?

Of course, we have already seen some of the things that our authors have to say about one or the other of these questions.

One class of reasons for denying that grade inflation is a problem would lie in putative evidence to the effect that either it does not exist, that it exists only in mild forms, or that we do not or cannot know that it exists. Adelman concedes that there has been an upward drift in grades in recent decades—from an average GPA of 2.66 for those who were twelfth graders in 1982 to 2.74 for those who were twelfth graders in 2000—but observes that this change is a minor one. Kamber denies that such a change is minor, partly because he understands the underlying problem differently. Such a rate of change would indeed be benign if it were economic inflation, but precisely because grade inflation differs from economic inflation such a trend, over time, can be seriously damaging, in his view. The conflation it causes remains in the system.

Depending on one's position on the underlying conceptual issue, one might well become very skeptical about whether we can know if there is any grade inflation at all in the long run. Kohn, Adelman, and Brighouse argue, based on the position that grade inflation is an upward shift in grades without a corresponding increase in achievement, that rising grades do not per se support the conclusion that there is inflation. In addition, Adelmam points out that, for a variety of reasons, we cannot compare what students achieve today with what their predecessors achieved in 1972, so that there is really no rational way to decide whether shifts in grading since then are inflationary or not. Brighouse gives a similar skeptical argument against the idea that there is grade *variation* between disciplines. Kamber, as we have seen, denies the definition of grade inflation on which such skeptical arguments are based. For him, an upward shift is inflationary if it decreases the system's ability to distinguish between levels of achievement—period!

Another class of reasons for thinking that grade inflation is not a problem would be one that supports the idea that *even if* it does exist, it does not cause any serious problems. Brighouse's "discounting" argument, which I have already mentioned, tends in this direction.

Finally, what, if anything, might be done about grade inflation? The one chapter in this volume that focuses on this issue is Kamber's "Combating Grade Inflation: Obstacles and Opportunities." The possible reforms that he considers are: (1) adopting caps that limit the

percent of a given grade that can be awarded (e.g., no more than 20% As in one class), (2) adjusting the value of submitted grades for a class upward when the class average is lower than a preferred average and downward when the class average is higher than a preferred average, (3) adding information on transcripts that clarifies or supplements submitted grades, and (4) requiring teachers to rank the students in each class. Of these four sorts of policies, the first one, grading caps, seems on the face of it to be the least politically feasible, mainly because it interferes most with what professors are allowed to do. Surprisingly, the faculty at Princeton have voted, and by a wide margin at that, to implement precisely this sort of policy. Kamber's recounting how this came about, and how it has been implemented, should be full of interest for would-be reformers at other institutions.

The most politically feasible type of response, Kamber thinks, is the third one: amended transcripts. Dartmouth and Columbia append to each grade on a transcript a note that indicates either the average grade for the class in which the grade was earned or the percentage of As given in the class. Such practices make grades more informative without interfering with what the instructor does, and so might be not be resisted by faculty as much as other reforms.

Brighouse discusses Valen Johnson's proposal that educational institutions deflate the higher grades of instructors who grade relatively laxly, replacing them with marks that are worth less (the second of Kamber's potential reforms, mentioned earlier). He rejects this on the grounds that it is politically infeasible, and because it would discourage the best students from taking courses from lax graders. Kamber points out (in "Combating Grade Inflation") that Johnson actually has a way to eliminate this effect, but it involves a sophisticated statistical formula that many faculty and students might find unintelligible.

Brighouse defends a reform that is regarded as radical in the United States. but is actually standard practice in Europe: separating the roles of teaching and evaluation. Teachers would continue to teach, but evaluation of student performance would be done by external examiners. By an interesting coincidence, Hunt describes what is essentially the same arrangement as the ideal way to evaluate the performance of *teachers*, of determining how successful they are at imparting knowledge. The only really effective way to judge teaching effectiveness would be to see how independent examiners would rate the proficiency of their students.

This volume concludes with a narrative by David T. Beito and Charles Nuckolls in which they tell of their attempts to bring about grading reform at the University of Alabama. Their campaign resulted in no change. It is instructive to compare their narrative with Kamber's narrative of reform attempts at Princeton, which did result in change. One difference that leaps forth is that the Princeton reform was in a sense a top-down movement, initiated by a dean, while the one at Alabama was a bottom-up one, initiated by two faculty members with no special institutional status. The Princeton dean was able to use her authority to focus attention on grading issues for eight years, before the crucial vote was taken. She formed a committee for the express purpose of doing so, and she chaired it herself. There are probably several lessons here that people who want to be agents of change would likely want to think about.

As I have said, I think many of those who thoughtfully read through this volume will find themselves in a state of confusion. But I think at least some of it would be a good sort of confusion, the sort that comes from a more nuanced understanding of the issues, and a grasp of the fundamental differences of principle that lie behind the positions that people take on those issues. As a small contribution to eliminating the not-so-good sorts of confusion, I will attempt, in a brief Afterword to this book, to say what, if anything, has been fairly definitely established by the preceding discussion.

Lester H. Hunt
Department of Philosophy
University of Wisconsin-Madison

NOTE

1. Harvey Mansfield, "Grade Inflation: It's Time to Face the Facts," *The Chronicle of Higher Education (*April 6, 2001).

ACKNOWLEDGMENTS

Thanks are due to the Wisconsin Association of Scholars (WAS) for a generous grant that made possible the academic conference from which most of the chapters in this book derive, and later that underwrote some of the expenses of producing this book. If it were not for the WAS and its dedication to the free exchange of ideas, then this volume would never have been possible. The Lynde and Harry Bradley Foundation also provided a substantial grant for the conference as well as this book; for that we are very grateful. I also would like to thank Deborah Katz Hunt for a great deal of crucial editorial work on this volume.

Two chapters in this book were originally published elsewhere. Alfie Kohn's "The Dangerous Myth of Grade Inflation" first appeared, in a slightly different form, in *The Chronicle of Higher Education* (November 8, 2002), p. B8. Francis K. Schrag's "From Here to Equality: Grading Practices for Egalitarians" first appeared, again minus some revisions, in *Educational Theory* (Winter 2001) 5:1. Reprinted by permission of the copyright holders.

This page intentionally left blank.

CHAPTER ONE

The Dangerous Myth of Grade Inflation

ALFIE KOHN

> Grade inflation got started . . . in the late '60s and early '70s. The grades that faculty members now give . . . deserve to be a scandal.
>
> *—Professor Harvey Mansfield, Harvard University, 2001*

> Grades A and B are sometimes given too readily—Grade A for work of no very high merit, and Grade B for work not far above mediocrity. . . . One of the chief obstacles to raising the standards of the degree is the readiness with which insincere students gain passable grades by sham work.
>
> *—Report of the Committee on Raising the Standard, Harvard University, 1894*

Complaints about grade inflation have been around for a very long time. Every so often a fresh flurry of publicity pushes the issue to the foreground again, the latest example being a series of articles in the *Boston Globe* in 2001 that disclosed—in a tone normally reserved for the discovery of entrenched corruption in state government—that a lot of students at Harvard were receiving As and being graduated with honors.

The fact that people were offering the same complaints more than a century ago puts the latest bout of harrumphing in perspective, not unlike those quotations about the disgraceful values of the younger generation that turn out to be hundreds of years old. The long history of indignation also pretty well derails any attempts to place the blame for higher grades on a residue of bleeding-heart liberal professors hired in the '60s. (Unless, of course, there was a similar countercultural phenomenon in the 1860s.)

Yet on campuses across America today, academe's usual requirements for supporting data and reasoned analysis have been suspended for some reason where this issue is concerned. It is largely accepted on faith that grade inflation—an upward shift in students' grade point averages without a similar rise in achievement—exists, and that it is a bad thing. Meanwhile, the truly substantive issues surrounding grades and motivation have been obscured or ignored.

The fact is that it is hard to substantiate even the simple claim that grades have been rising. Depending on the time period we are talking about, that claim may well be false. In their book *When Hope and Fear Collide*, Arthur Levine and Jeanette Cureton tell us that more undergraduates in 1993 reported receiving As (and fewer reported receiving grades of C or below) compared to their counterparts in 1969 and 1976 surveys.[1] Unfortunately, self-reports are notoriously unreliable, and the numbers become even more dubious when only a self-selected, and possibly an unrepresentative, segment bothers to return the questionnaires. (One out of three failed to do so in 1993; no information is offered about the return rates in the earlier surveys.)

To get a more accurate picture of whether grades have changed over the years, one needs to look at official student transcripts. Clifford Adelman, a senior research analyst with the U.S. Department of Education, did just that, reviewing transcripts from more than 3,000 institutions and reporting his results in 1995. His finding: "Contrary to the widespread lamentations, grades actually declined slightly in the last two decades." Moreover, a report released in 2002 by the National Center for Education Statistics revealed that fully 33.5 percent of American undergraduates had a grade point average of C or below in 1999–2000, a number that ought to quiet "all the furor over grade inflation," according to a spokesperson for the Association of American Colleges and Universities. (A review of other research suggests a comparable lack of support for claims of grade inflation at the high school level.)[2]

However, even where grades *are* higher now compared to then, that does not constitute proof that they are inflated. The burden rests with critics to demonstrate that those higher grades are undeserved, and one can cite any number of alternative explanations. Maybe students are turning in better assignments. Maybe instructors used to be too stingy with their marks and have become more reasonable. Maybe the concept of assessment itself has evolved, so that today it is more a means for allowing students to demonstrate what they know rather

than for sorting them or "catching them out." (The real question, then, is why we spent so many years trying to make good students look bad.) Maybe students are not forced to take as many courses outside their primary areas of interest in which they did not fare as well. Maybe struggling students are now able to withdraw from a course before a poor grade appears on their transcripts. (Say what you will about that practice, it challenges the hypothesis that the grades students receive in the courses they complete are inflated.)

The bottom line: No one has ever demonstrated that students today get As for the same work that used to receive Bs or Cs. We simply do not have the data to support such a claim. Consider the most recent, determined effort by a serious source to prove that grades are inflated: "Evaluation and the Academy: Are We Doing the Right Thing?," a report released in 2002 by the American Academy of Arts and Sciences.[3] Its senior author is Henry Rosovsky, formerly Harvard's dean of the faculty. The first argument offered in support of the proposition that students could not possibly deserve higher grades is that SAT (Scholastic Aptitude Test) scores have dropped during the same period that grades are supposed to have risen. But this is a patently inapt comparison, if only because the SAT is deeply flawed. It has never been much good even at predicting grades during the freshman year in college, to say nothing of more important academic outcomes. A four-year analysis of almost 78,000 University of California (UC) students, published in 2001 by the UC president's office, found that the test predicted only 13.3 percent of variation in freshman grades, a figure roughly consistent with hundreds of previous studies.[4]

Even if one believes that the SAT is a valid and valuable exam, however, the claim that scores are dropping is a poor basis for the assertion that grades are too high. First, it is difficult to argue that a standardized test taken in high school and grades for college course work are measuring the same thing.

Second, changes in aggregate SAT scores mostly reflect the proportion of the eligible population that has chosen to take the test. The American Academy's report states that average SAT scores dropped slightly from 1969 to 1993. But over that period, the pool of test takers grew from about one-third to more than two-fifths of high school graduates—an addition of more than 200,000 students.

Third, a decline in overall SAT scores is hardly the right benchmark against which to measure the grades earned at Harvard or other elite

institutions. Every bit of evidence I could find—including a review of the SAT scores of entering students at Harvard over the past two decades, at the nation's most selective colleges over three and even four decades, and at all private colleges since 1985—uniformly confirms a virtually linear rise in both verbal and math scores, even after correcting for the renorming of the test in the mid-1990s. To cite just one example, the latest edition of "Trends in College Admissions" reports that the average verbal-SAT score of students enrolled in all private colleges rose from 543 in 1985 to 558 in 1999. Thus those who regard SAT results as a basis for comparison should *expect* to see higher grades now rather than assume that they are inflated.

The other two arguments made by the authors of the American Academy's report rely on a similar sleight of hand. They note that more college students are now forced to take remedial courses but offer no reason to think that this is especially true of the relevant student population—namely, those at the most selective colleges who are now receiving As instead of Bs.[5]

Finally, they report that more states are adding high school graduation tests and even standardized exams for admission to public universities. Yet that trend can be explained by political factors and offers no evidence of an objective decline in students' proficiency. For instance, scores on the National Assessment of Educational Progress, known as "the nation's report card" on elementary and secondary schooling, have shown very little change over the past couple of decades, and most of the change that has occurred has been for the better. As David Berliner and Bruce Biddle put it in their tellingly titled book *The Manufactured Crisis*,[6] the data demonstrate that "today's students are at least as well informed as students in previous generations." The latest round of public school bashing—and concomitant reliance on high-stakes testing—began with the Reagan administration's "Nation at Risk" report, featuring claims now widely viewed by researchers as exaggerated and misleading.

Beyond the absence of good evidence, the debate over grade inflation brings up knotty epistemological problems. To say that grades are not merely rising but inflated—and that they are consequently "less accurate" now, as the American Academy's report puts it — is to postulate the existence of an objectively correct evaluation of what a student (or an essay) deserves, the true grade that ought to be uncovered and

honestly reported. It would be an understatement to say that this reflects a simplistic and an outdated view of knowledge and learning.

In fact, what is most remarkable is how rarely learning even figures into the discussion. The dominant disciplinary sensibility in commentaries on this topic is not that of education—an exploration of pedagogy or assessment—but rather of economics. That is clear from the very term "grade inflation," which is, of course, just a metaphor. Our understanding is necessarily limited if we confine ourselves to the vocabulary of inputs and outputs, incentives, resource distribution, and compensation.

Suppose, for the sake of the argument, we assumed the very worst —not only that students are getting better grades than did their counterparts of an earlier generation, but that the grades are too high. What does that mean, and why does it upset some people so?

To understand grade inflation in its proper context, we must acknowledge a truth that is rarely named: The crusade against it is led by conservative individuals and organizations who regard it as analogous—or even related—to such favorite whipping boys as multicultural education, the alleged radicalism of academe, "political correctness" (a label that permits the denigration of anything one does not like without having to offer a reasoned objection), and too much concern about students' self-esteem. Mainstream media outlets and college administrators have allowed themselves to be put on the defensive by accusations about grade inflation, as can be witnessed when deans at Harvard plead nolo contendere and dutifully tighten their grading policies.

What are the critics assuming about the nature of students' motivation to learn, about the purpose of evaluation and of education itself? (It is surely revealing when someone reserves time and energy to complain bitterly about how many students are getting As—as opposed to expressing concern about, say, how many students have been trained to think that the point of going to school is to get As.)

"In a healthy university, it would not be necessary to say what is wrong with grade inflation," Harvey Mansfield asserted in an opinion article in 2001 in the *Chronicle of Higher Education.*[7] That, to put it gently, is a novel view of health. It seems reasonable to expect those making an argument to be prepared to defend it, and also valuable to bring their hidden premises to light. The assumptions that follow seem to underlie the grave warnings about grade inflation:

The professor's job is to sort students for employers or graduate schools. Some are disturbed by grade inflation—or, more accurately, grade compression—because it then becomes harder to spread out students on a continuum, ranking them against one another for the benefit of postcollege constituencies. One professor asks, by way of analogy, "Why would anyone subscribe to *Consumers Digest* if every blender were rated a 'best buy'?"

But how appropriate is such a marketplace analogy? Is the professor's job to rate students like blenders for the convenience of corporations, or to offer feedback that will help students learn more skillfully and enthusiastically? (Notice, moreover, that even consumer magazines do not grade on a curve. They report the happy news if it turns out that every blender meets a reasonable set of performance criteria.)

Furthermore, the student-as-appliance approach assumes that grades provide useful information to those postcollege constituencies. Yet growing evidence—most recently in the fields of medicine and law, as cited in publications such as the *Journal of the American Medical Association* and the *American Educational Research Journal*—suggests that grades and test scores do not in fact predict career success, or much of anything beyond subsequent grades and test scores.

Students should be set against one another in a race for artificially scarce rewards. "The essence of grading is exclusiveness," Mansfield said in one interview. Students "should have to compete with each other," he said in another. A chemistry professor at the University of Wisconsin at La Crosse said, "We cannot have half our students at the head of the class." In other words, even when no graduate school admissions committee pushes for students to be sorted, they ought to be sorted anyway, with grades reflecting relative standing rather than absolute accomplishment. In effect, this means that the game should be rigged so that no matter how well students do, only a few can get As. The question guiding evaluation in such a classroom is not "How well are they learning?" but "Who is beating whom?" The ultimate purpose of good colleges, this view holds, is not to maximize success but to ensure that there will always be losers.

A bell curve may sometimes—but only sometimes—describe the range of knowledge in a roomful of students at the beginning of a course. When it is over, though, any responsible educator hopes that the results would skew drastically to the right, meaning that most students learned what they had not known before. Thus in their important

study *Making Sense of College Grades*, Ohmer Milton, Howard Pollio, and James Eison write, "It is not a symbol of rigor to have grades fall into a 'normal' distribution; rather, it is a symbol of failure—failure to teach well, failure to test well, and failure to have any influence at all on the intellectual lives of students."[8] Making sure that students are continually re-sorted, with excellence turned into an artificially scarce commodity, is almost perverse. Excellence is not the same thing as victory.

What does relative success signal about student performance in any case? The number of peers that a student has bested tells us little about how much she knows and is able to do. Moreover, such grading policies may create a competitive climate that is counterproductive for winners and losers alike, to the extent that it discourages a free exchange of ideas and a sense of community that is conducive to exploration.

Harder is better (or higher grades mean lower standards). Compounding the tendency to confuse excellence with victory is a tendency to confuse quality with difficulty—as evidenced in the accountability fad that has elementary and secondary education in its grip just now, with relentless talk of "rigor" and "raising the bar." The same confusion shows up in higher education when professors pride themselves not on the intellectual depth and value of their classes but merely on how much reading they assign, how hard their tests are, how rarely they award good grades, and so on. "You're going to have to *work* in here!" they announce, with more than a hint of machismo and self-congratulation.

Some people might defend that posture on the grounds that students will perform better if As are harder to come by. In fact, the evidence on this question is decidedly mixed. Stringent grading sometimes has been shown to boost short-term retention as measured by multiple-choice exams—never to improve understanding or promote interest in learning. An analysis, released in 2000 by Julian R. Betts and Jeff Grogger, professors of economics at the University of California at San Diego and at Los Angeles, respectively, found that tougher grading was initially correlated with higher test scores. But the long-term effects were negligible—with the exception of minority students, for whom the effects were negative.

It appears that something more than an empirical hypothesis is behind the "harder is better" credo, particularly when it is set up as a painfully false dichotomy: Those easy-grading professors are too lazy to care, or too worried about how students will evaluate them, or overly concerned about their students' self-esteem, whereas *we* are the last

defenders of what used to matter in the good old days. High standards! Intellectual honesty! No free lunch!

The American Academy's report laments an absence of "candor" about this issue. Let us be candid then. Those who grumble about undeserved grades sometimes exude a cranky impatience with—or even contempt for—the late adolescents and young adults who sit in their classrooms. Many people teaching in higher education, after all, see themselves primarily as researchers and regard teaching as an occupational hazard, something they are not very good at, were never trained for, and would rather avoid. It would be interesting to examine the correlation between one's view of teaching (or of students) and the intensity of one's feelings about grade inflation. Someone also might want to examine the personality profiles of those who become infuriated over the possibility that someone, somewhere, got an A without having earned it.

Grades motivate. With the exception of orthodox behaviorists, psychologists have come to realize that people can exhibit qualitatively different kinds of motivation: intrinsic, in which the task itself is seen as valuable, and extrinsic, in which the task is just a means to the end of gaining a reward or escaping a punishment. The two are not only distinct but often inversely related. Scores of studies have demonstrated, for example, that the more people are rewarded, the more they come to lose interest in whatever had to be done in order to get the reward. (That conclusion is essentially reaffirmed by the latest major meta-analysis on the topic: a review of 128 studies, published in 1999 by Edward L. Deci, Richard Koestner, and Richard M. Ryan.)[9]

Those unfamiliar with that basic distinction, let alone the supporting research, may be forgiven for pondering how to "motivate" students, then concluding that grades are often a good way of doing so, and consequently worrying about the impact of inflated grades. But the reality is that it does not matter how motivated students are; what matters is *how* students are motivated. A focus on grades creates, or at least perpetuates, an extrinsic orientation that is likely to undermine the love of learning we are presumably seeking to promote.

Three robust findings emerge from the empirical literature on the subject: Students who are given grades, or for whom grades are made particularly salient, tend to display less interest in what they are doing, fare worse on meaningful measures of learning, and avoid more challenging tasks when given the opportunity—as compared to those in a

nongraded comparison group. College instructors cannot help noticing, and presumably being disturbed by, such consequences, but they may lapse into blaming students ("grade grubbers") rather than understanding the systemic sources of the problem. A focus on whether too many students are getting As suggests a tacit endorsement of grades that predictably produces just such a mind-set in students.

These fundamental questions are almost completely absent from discussions of grade inflation. The American Academy's report takes exactly one sentence—with no citations—to dismiss the argument that "lowering the anxiety over grades leads to better learning," ignoring the fact that much more is involved than anxiety. It is a matter of why a student learns, not only how much stress he feels. Nor is the point just that low grades hurt some students' feelings, but that grades, per se, hurt all students' engagement with learning. The meaningful contrast is not between an A and a B or C but between an extrinsic and an intrinsic focus.

Precisely because that is true, a reconsideration of grade inflation leads us to explore alternatives to our (often unreflective) use of grades. Narrative comments and other ways by which faculty members can communicate their evaluations can be far more informative than letter or number grades, and much less destructive. Indeed, some colleges —for example, Hampshire, Evergreen State, Alverno, and New College of Florida—have eliminated grades entirely, as a critical step toward *raising* intellectual standards. Even the American Academy's report acknowledges that "relatively undifferentiated course grading has been a traditional practice in many graduate schools for a very long time." Has that policy produced lower-quality teaching and learning? Quite the contrary: Many people say they did not begin to explore ideas deeply and passionately until graduate school began and the importance of grades diminished significantly.

If the continued use of grades rests on nothing more than tradition ("We've always done it that way"), a faulty understanding of motivation, or excessive deference to graduate school admissions committees, then it may be time to balance those factors against the demonstrated harms of getting students to chase As. Milton and his colleagues discovered—and others have confirmed—that a "grade orientation" and a "learning orientation" on the part of students tend to be inversely related. That raises the disturbing possibility that some colleges are institutions of higher learning in name only, because the paramount

question for students is not "What does this mean?" but "Do we have to know this?"

A grade-oriented student body is an invitation for the administration and faculty to ask hard questions: What unexamined assumptions keep traditional grading in place? What forms of assessment might be less destructive? How can professors minimize the salience of grades in their classrooms so long as grades must still be given? And, if the artificial inducement of grades disappeared, then what sort of teaching strategies might elicit authentic interest in a course?

To engage in this sort of inquiry, to observe real classrooms, and to review the relevant research is to arrive at one overriding conclusion: The real threat to excellence is not grade inflation at all; it is grades.[10]

NOTES

1. Arthur Levine and Jeanette Cureton, *When Hope and Fear Collide* (San Francisco: Jossey-Bass, 1998).

2. A subsequent analysis by Adelman, which reviewed college transcripts from students who were graduated from high school in 1972, 1982, and 1992, confirmed that there was no significant or linear increase in average grades over that period. The average GPA for those three cohorts was 2.70, 2.66, and 2.74, respectively. The proportion of As and Bs received by students was 58.5 percent in the '70s, 58.9 percent in the '80s, and 58.0 percent in the '90s. Even when Adelman looked at "highly selective" institutions, he again found very little change in average GPA over the decades.

3. Henry Rosovsky and Matthew Hartley, "Evaluation and the Academy: Are We Doing the Right Thing? Grade Inflation and Letters of Recommendation," Occasional Paper (Cambridge, MA: American Academy of Arts and Sciences, 2002).

4. I outlined numerous other problems with the test in "Two Cheers for an End to the SAT," *The Chronicle,* March 9, 2001, available at www.alfiekohn.org.

5. Adelman's newer data challenge the premise that there has been *any* increase. In fact, "the proportion of all students who took at least one remedial course (in college) dropped from 51 percent in the [high school] class of 1982 to 42 percent in the class of 1992."

6. David Berliner and Bruce Biddle, *The Manufactured Crisis* (Reading, MA.: Addison-Wesley, 1995).

7. Harvey Mansfield, "Grade Inflation: It's Time to Face the Facts," *Chronicle of Higher Education*, April 6, 2001.

8. Ohmer Milton, Howard Pollio, and James Eison, *Making Sense of College Grades* (San Francisco: Jossey-Bass, 1986).

9. Edward L. Deci, Richard Koestner and Richard M. Ryan, "A Meta-analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation," *Psychological Bulletin* 125, (1999): 627–68.

10. This article is based on one that first appeared in *Chronicle of Higher Education* 49:11 (November 8, 2002) and Mr. Kohn's talk at the "Academic Standards and Grade Inflation Conference," October 11, 2003.

This page intentionally left blank.

CHAPTER TWO

Undergraduate Grades: A More Complex Story Than "Inflation"

CLIFFORD ADELMAN

Some decades ago, in teaching freshman composition, I offered students a small canon of advisements, one of which was never to use the demonstrative adjective "this" in exposition without attaching it to a noun, hence rendering it a descriptive adjective. Tell your reader "this *what*," and you will find your reader unconsciously grateful.

When it comes to something called "grade inflation," it is difficult to follow this advice. I do not know what the "this" is, and when one reads the modest volume of academic journal literature (much of it dated), and the more considerable volume of screeds, polemics, op-eds, and "fugitive" pieces parked on uniform resources locators (URLs), there is not much to help me out. With the exception of some institutional studies (e.g., Beck 1999; McSpirit and Jones 1999; Olsen 1997; Smith 1992), along with research on what aspects of student academic work behaviors make a difference in grades (e.g., Michaels and Miethe 1989; Farkas and Hotchkiss 1989), and some of the empirical work of economists (e.g., Freeman 1999), adequate evidence and appropriate statistical methodology seem disposable whenever the topic arises. A topic such as this (whatever "this" is) deserves systematic treatment (for an exemplary case, despite its reliance on student self-reported grades, see Kuh and Hu 1999).

This contribution to this collection of chapters is intended to place some large-scale, national time series evidence on the table, to reflect on what it might indicate (what the "this" means), and to suggest some future topics and lines of research.

The topic sentences of this contribution are fairly simple:

- "Inflation" in the judgment of human intellectual performance in higher education contexts cannot be proven—one way or the other. We do not have the tools, and there are no convincing economic analogues.
- When we turn to proxy measures of change in the judgment of human intellectual performance in higher education and use the *student* as the unit of analysis, national transcript-based data for traditional-age cohorts during the period 1972–2000 do not support contentions of across-the-board linear acceleration of average GPAs. The most we can say, in terms of effect size, is that there has been a small-to-moderate increase in the average GPA of those who earned the bachelor's degree.
- When we use the *course* as the unit of analysis and examine the distribution of letter grades by institutional selectivity, the national transcript-based data for traditional-age cohorts do not support contentions of increasing proportion (let alone increasing dominance) of honors grades (As, in particular), no matter what degree of institutional selectivity is at issue.
- On the other hand, the national time series data introduced in the modern watershed year for U.S. higher education, 1972, reveal an increasing proportion of letter grades *removed from* the calculation of GPA altogether: "P" (or "CR") in pass/fail courses, "W" (for nonpenalty withdrawals, as distinct from drops), and "NCR" (the data sets' abbreviation for no-credit-repeats). These are cases in which the judgment of human intellectual performance is avoided. We can call this "grade devaluation" because it reduces the importance of the signaling features of the residual markers of the traditional grading system.
- The tapestry signaling the quality of student academic performance includes public recognitions such as Phi Beta Kappa and graduation with honors, and less public notices of dean's lists, academic probation and academic dismissal. These compressed signals transcend GPAs and letter-grade distributions.
- Whatever problems are perceived to exist in grading are best addressed at the institutional level, and by systematic construction of criterion-referenced high-stakes assessments that can be repeated with a high degree of reliability at regular intervals. This is not an easy task.

The large-scale evidence comes from the postsecondary transcripts of three national longitudinal studies conducted by the National Center for Education Statistics of the U.S. Department of Education. These studies followed members of the scheduled high school graduating classes of 1972, 1982, and 1992[1] for a minimum of twelve years. For those students who entered the postsecondary system, transcripts were collected at the end of the study period, and with a response rate of over 90 percent from over 2,500 institutions of all kinds (research universities, community colleges, trade schools). Whatever the occasional problems with transcripts, they neither lie, exaggerate, nor forget, particularly when gathered in a way that does not allow anyone to screen out undesirable cases. The data they present are more reliable and valid than those derived from the questionable source of student self-reported grades[2] one finds in Levine and Cureton (1998) or Kuh and Hu (1999). The national transcript archives are unobtrusive data, and in discussions of grades and grading in higher education, we ignore them at our willful peril.

Other data on grades are available for public examination, though they are summary and secondary reports, not primary sources. What distinguishes the data of the national longitudinal studies from those assembled from eighty-three institutions listed on gradeinflation.com in July 2005, for example, is that the national studies cover the *same* cohort-representative[3] populations in the *same* time periods with a standardized grade coding system on the *same* scale and with the *same* source–a transcript—and present GPAs with standard errors or standard deviations that allow the true judgment of change over time by z-scores, standard deviation units, or effect size. All of this information is available on CD-ROM, and the construction of all variables is detailed in descriptive windows. With a license to use the restricted files (not difficult to obtain) analysts can create their own parallel universes. Nothing is hidden.[4]

In contrast, the eighty-three institutions on gradeinflation.com as of July 2005 present seventy-nine different reporting periods, ranging from four years to forty years. In less than thirty of those eighty-three cases can the data be traced to an unassailable source—an institutional research office or a registrar's office (other sources include local newspapers, student newspapers, the national trade press and in one case, a report from school A that tattles on school B). The

gatekeeper for gradeinflation.com says only one school reported a declining average undergraduate GPA, but when one clicks through links to the underlying reports and examines the rest of the data, there are seven prima facie cases of no change and another dozen in which change is doubtful. For example, increases in the average GPA from 2.90 to 2.96 over the 1997–2000 period, or from 2.80 to 2.86 over a decade, or 3.33 to 3.38 from 1997 to 2002, or 3.16 to 3.28 over thirty years are meaningless even if they are statistically significant (and I doubt they would prove to be statistically significant). It also is unclear whether the reported data are for degree recipients only or everybody-in-progress, or whether local changes in grading symbols confound the results (Millman, Slovacek, Kulick, and Mitchell 1983). Some of the reports prepared by institutional research offices (the most credible of the sources) distinguish between freshman grades and senior grades, for example. Some distinguish between full-time and part-time students. But for a significant number of these reports, the identity of the student universe is a mystery. One thus emerges with little sense of what all these numbers are really measuring.

WHAT IS THE "THIS"? MEDIA DYNAMICS AND SYMBOLS

Part of "this" is about the origins, media, and distribution of a story from private to public arenas. It is a communications saga. We did not get it from the academic journals; we got it from the newspapers, the op-eds, and the trade press.[5] Of every three stories that appear in the general and trade press on this issue, two are situated in institutions such as Princeton and Amherst, the type of school that controls the nature and flow of information about higher education to the major media distribution nodes. The media, in turn, are delighted to report that a higher percentage of grades awarded in courses at Midas-touch colleges are "As" than was the case twenty or twenty-five or thirty years ago. Even if less than 2 percent of undergraduates, and about 5 percent of bachelor's recipients attend such institutions, the distribution of grades at those places, like the graduation rates of NCAA Division I varsity football and basketball players, is a glitzy enough headline to convey the impression that something equivalent to a plague infests the entire system. Three data anecdotes from Mt. Sinai start a trend that is ultimately joined by watercooler stories from

institutions with shaky accreditation credentials. All of it conveys the impression that wherever one turns in the vast enterprise of U.S. higher education, the judgment of human performance, indicated by traditional proxy symbols, is, at best, lax.

When the topic is conveyed through academic and scholarly journals, the main line of the story emanates from topic sentences such as "In the second half of the twentieth century, grade inflation has become an embarrassing fact of academic life in the United States," even when the authors continue, "Although research is sparse, there is evidence that the inflation exists" (Bearden, Wolfe, and Grosch 1992, 745). Whether from the general and trade press or the journals, most writing on the topic employs ostensive definition: it observes GPAs rising, the percentage of As rising, or, in one somewhat dated study (Bromley, Crow, and Gibson 1978), a rising percentage of students graduating with honors,[6] points to the phenomenon, and sentences it to "inflation." Grade inflation becomes a given, analogous to a theorem in geometry. The writer states it as a fact, then explains, rather tersely and opaquely, what it is before turning to an account of how we reached this sorry state. For example:

> When a grade is viewed as less rigorous than it ought to be. (Milton, Pollio, and Eison, p. 29, 1986)
>
> . . .an upward shift in the grade point average (GPA) over an extended period of time without a corresponding increase in student achievement." (Rosovsky and Hartley 2002, and others, p. 4)
>
> . . .the lowering in value of As and Bs because of their more frequent use." (Millman, Slovacek, Kulick, and Mitchell 1983, p. 423)

Given the reference points in the ostensive definitions of "grade inflation," part of the "this" is informed by semiotics and the ways established signs get separated from the realities we assume they represent. People who write about grades and grading assume a closed universe of symbols: A, B, C, D, E, F, P, S, U, sometimes with pluses and minuses, sometimes in mixtures such as BC. On college transcripts, however, one finds a much wider range of signs, reflecting the ways

in which the alphabetic symbol system is utilized more robustly to mark different forms of student behavior or emerging institutional practices. In both the polemics and most of the research about grades, these other symbols are never accounted for, though, as we will see, they may be more important.

In the course of coding the 60,000 pieces of paper in the most recent national transcript sample, covering the period 1992–2000, we came across grades of X, M, Z, CR, NC, RP, WP, WF, IP, DE, EI, NW, and others. Some of these are obvious (WP, WF, CR, NC, IP); others were explained by the guides institutions provide for the interpretation of transcripts; but others required telephone calls to registrars. One of our tasks in data entry and editing across roughly 400,000 discrete course entries from 2,500 to 3,000 institutions in each of the three transcript files was to standardize grades[7] with (1) a recognizable, bounded letter system that (2) could be converted into a 4-point scale with a limited number of reserved codes (e.g., –4 for audits [AU], incompletes [I], and in-progress [IP] courses that would not be included in any aggregates of credits or numerical grades).

Recalling Birnbaum's (1977) point that grades are not grade point averages (the formulas for which differ from institution to institution), the data decision rules also standardized the calculation of GPA across all institutions represented. No-penalty withdrawals (as distinguished from drops), no credit courses, nonadditive credit remedial courses, and the first taking of a course that is repeated for a higher grade are not included in the calculation of GPA. All these procedures ultimately have the student in mind as the unit of analysis, no matter how many schools the student attends (57 percent in the 1992–2000 sample attended more than one institution as undergraduates; 22 percent attended more than two). The student's cumulative GPA (at the analysts' choice of three points in time[8]) is set on a percentile scale. So it is perfectly possible for a graduating senior with a 3.4 average gleaned from three different institutions to wind up in the 71st percentile of a nationally representative population. The percentile scale could as easily become the dependent variable in a linear analysis as the 4-point scale; in a logistic analysis, the 71st percentile student might fall on one side of a dichotomous variable marking those in the top 40 percent of GPA. In these analytic modes, the conventional signs—alphabetic, 4-point, or numerical—are moot.

WHAT IS THE "THIS"? PRICE, VALUE, AND CONSUMER ECONOMICS

Much of the "this" is about economics. McKenzie and Staaf (1974) argue that when faculty raise their grade distributions, they lower the price of time that students must invest to earn an honors grade, thus freeing the student for more leisure which, McKenzie and Staaf assume, was the student's primary market objective. Marginal students in particular, they claim, are attracted to courses where the odds of a higher grade for less effort are known. Highly talented students, they admit, do not behave according to classic economic laws. This economic approach shifts easily to the academic department (or individual faculty member) in terms of the production function, under the assumption that the product of academic departments is enrollment volume, and the way to get there is to lower the price, that is, raise the grade distribution in relation to effort (Sabot and Wakeman-Linn 1991; Stone 1995), or, as Freeman (1999) demonstrates, to raise the distribution in inverse relationship to the expected income of graduates majoring in the field at issue.

McKenzie and Staaf's odd approach treats grades as commodities for which students pay in time. This chapter holds the commonsense contrary: the "commodity" is an artifact of student production—a paper, a test, a project, a laboratory, a performance—and/or a collection of such productions within a course. The "price" is the grade that which, the faculty member is willing to "pay" (assign) for the particular artifact, performance, or collections of products at issue. If one purports to talk about something called "inflation," then one normally references change in price. In the ideal determination of something we call "inflation" in daily life, we encounter one or more of the following phenomena:

- The price of a given product or commodity or service of *constant* composition or characteristics rises faster than does our income, which in turn reflects the general money supply. We judge the nominal prices of shaving cream and shampoo, subway tokens, haircuts, and, of course, gasoline, this way.
- The price of a given product or commodity or service remains stable while the perceived or measured quality of the product, commodity, or service declines.

- The price of a given product or commodity declines, but at a slower rate than either a decline in our income or a decline in perceived or measured quality.
- The price of a given product or commodity rises faster than does the perceived or measured quality of the same product or commodity, or, that the price rises while the quality falls.

These permutations of the relationship of price to quality are what one might call "commodity views" of inflation. If the judgment of student academic performance met any of the criteria for "inflation" as we think of it in the dailiness of our lives, then it might be a matter worth noting. But to prove that something analogous to inflation exists in the judgment of human intellectual performance requires the *same* type of assessments with the *same* prompts based on the *same* material judged by panels observing the *same* criteria (and with proven high inter-rater reliability)—all in two or more periods of time[9].

The basic conditions for this framework of judgment do not exist in higher education. Birnbaum (1977) struggled with this void in the context of grades as proxy measures and came up with a definition that (perhaps unintentionally) underscores the problem:

> . . .inflation can be viewed as a process in which a defined level of academic achievement results in a higher grade than awarded to that level of academic achievement in the past. (522)

The key phrase is "a defined level of academic achievement" and its accompanying assumption that the "definition" remains constant. Olsen (1997) indirectly raises a similar reference point when he asserts that "inflation is . . . the state of being expanded to an abnormal or unjustifiable level" (6). As applied to the judgment of human intellectual performance, that raises the question of what is "normal" let alone what is "justifiable" (a different kind of judgment). When one thinks of our examples in the commodity view of inflation, the question of how one marks a "normal" or "justifiable" price for shampoo, gasoline, or haircuts is one that is answered (if at all) in a nexus of price history, changes in personal income, and geographical region and urbanicity of the point of sale, among other factors. These conditions do not translate to higher education.

It is not merely the case that the basic conditions for the commodity view of inflation—those that allow us to determine what is "normal," those that assume a constant "defined level of academic achievement," those that allow a reliable statement of "justifiability"—do not exist in higher education. In a very important sense, these conditions *should not* exist. If we are teaching the same material and giving the same tests—with the same criteria for judgment, the same "defined level of academic achievement"—that we gave twenty years ago, then we are not fulfilling the principal purposes for which students come to us or the expectations that communities and economies have of us. Furthermore, there are few uniform commodities in higher education, and new fields of study, with distinct models of assessment, are constantly arising–as they should, and as we expect in a universe in which knowledge rarely sits still long enough to be measured.

PRICES AND PROXIES

Virtually all writers on the topic of grade inflation, whether disciplined researchers or less disciplined polemicists, implicitly acknowledge that we cannot apply the commodity view of inflation to the judgment of human intellectual performance in a rigorous manner. Instead, they turn to double proxy indicators. The first proxy is the trend in grades. The second is the trend in standardized test scores of entering students, a proxy for putative ability of the student to turn latent talent into performances deserving of the grades in question.

The test scores are de facto prices of other standard educational commodities, and the writers assume there must be grade inflation because GPAs are rising/there are more As while SAT and/or ACT (American College Test) scores are stable, declining, or rising by a percentage less than GPAs are rising (Prather, Smith, and Kodras 1979; Rosovsky and Hartley 2002; Birnbaum 1977).

This definition of "inflation" has more faults than San Andreas. First, and most important, the SAT and ACT are measures of general learned abilities, while grades in microbiology, Japanese history of the Tokugawa period, or cost accounting judge student mastery of very specific disciplinary material. To compare one to the other is analogous to saying that my judgment of the price of an automobile relative to its quality is based on the change (guaranteed!) in the price of gasoline. Yes, they are related, but distantly.

Second, if one contends that measure 1 has changed more/faster than measure 2, then, to put it politely, it would be helpful if one converted them to the same metric, for example, z-score, standard deviation units, or effect size, and compared them in terms of those metrics.

Third, grades are assigned for all enrolled undergraduates, including "nontraditional" students who entered higher education for the first time years after they graduated from high school and may/may not have generated the SAT or ACT scores to which the comparison refers. In the Beginning Postsecondary Students Longitudinal Study of 1995/96–2001, some 5 percent of beginning students at four-year colleges in 1995–96, 28 percent at community colleges, and 41 percent at other sub-baccalaureate institutions were at least twenty-four years old. Of those age twenty-four or older, 95 percent had never taken either the SAT or ACT.[10] Unless the analysis is sophisticated enough to account for these disconnects, it is highly suspect, to say the least.

If this analysis were post hoc, and referenced subtest scores on GRE (Graduate Record Examination) field examinations, linking them to grades in matching undergraduate course work, for example, the Computational Mathematics section of the Computer Science field test and grades in courses such as combinatorics, numerical methods, and/or linear algebra (Oltman 1982), then it might be more convincing.[11] Smith (1992) postulated a more elaborate external criterion configuration of examinations, including the GRE field tests, and found "positive and often strong correlations" between grades and performance on the examinations. While Smith's study was confined to one institution, it suggests that the very approach to external criteria—the use of second proxy measures—(1) needs radical overhaul), and (2) warrants extension to a representative sample of other institutions to test its generalizability.

WHAT THE TRANSCRIPT DATA SAY

If the relationship between the proxy measures of postsecondary grade distribution trends and GPAs, on the one hand, and pre-collegiate test scores as a putatively stable reference point of student ability or achievement is dubious, and if this relationship is nonetheless the default of the inflation argument, a closer examination of the proxy measures alone is warranted. Let's look at the national transcript-based data

sets, what they tell us, and think about ways that serious research might use them in the future.

Table 2.1 scans across the transcript-recorded postsecondary grades of the three completed national grade-cohort longitudinal studies using two common metrics, the distribution of letter grades and GPAs. In Table 2.1, the student is the unit of analysis. Only undergraduate grades are included.

The first—and major—point is that, judging by both distribution of letter grades and GPAs, changes have been minor and complex since the high school class of 1972 went to college. In terms of the distribution of letter grades, the proportion of grades that were "A" declined slightly between the Class of 1972 and the Class of 1982, then rose between the Class of 1982 and the Class of 1992. The inverse to this pattern can be observed for the proportions of grades that were "B" and "D." In terms of final undergraduate GPAs, those for women and students who earned bachelor's degrees, as well as among some majors (health sciences and services, social sciences, and applied social sciences), dropped from the Class of 1972 to the Class of 1982, then rose for the Class of 1992.

Think first about the stable-to-downward slope of GPAs between the 1970s and 1980s.[12] This slope during a period of massification of the higher education system in this country is a matter of common sense. It is what one would expect; it is what we got; and it has been marked by others (e.g., Summerville and Ridley 1990). Between 1976 and 1980, women increased their undergraduate enrollment by 21 percent (versus 2 percent for men) and had become a majority of undergraduates. Women continued on this trajectory, increasing their undergraduate enrollment by another 20 percent (versus 8 percent for men) between 1980 and 1990 (Snyder 2004, table 209, p. 258). While women have historically earned higher grades than men, when the pool expanded by as much as it did in the 1980s, women's average GPA inevitably fell (following statistical conventional wisdom), though it still remained higher than men's average GPA.

Then turn to the generally upward slope in GPAs between the 1980s and 1990s, and consider the letter grade distribution in terms of the increase in the percentage of grades removed from GPA calculations (the Pass/Credit line, and the Withdrawal/Repeat line). Roughly 5.5 percent of the standard letter grades of the 1980s became either

Table 2.1
Distribution of Undergraduate Letter Grades of 1972, 1982, and 1992 Twelfth Graders, and Average Undergraduate Grade Point Averages (GPAs) of 1972, 1982, and 1992 Twelfth Graders by Gender, Level of Educational Attainment, and (for Bachelor's Degree Recipients), bachelor's degree major

	Class of 1972	Class of 1982	Class of 1992
	Distribution of Letter-Equivalent Grades[1]		
As	27.3 (0.34)	26.1 (0.33)	28.1 (0.36)
Bs	31.2 (0.24)	32.8 (0.27)	29.9 (0.23)
Cs	21.9 (0.21)	22.2 (0.23)	18.2 (0.23)
Ds	5.4 (0.14)	5.8 (0.12)	4.6 (0.09)
Fs/penalty grades	3.8 (0.11)	4.8 (0.13)	4.5 (0.14)
Pass/credit, etc.[2]	6.4 (0.15)	2.6 (0.17)	6.4 (0.17)
Withdrawal, no-credit repeat[2]	4.0 (0.13)	6.7 (0.16)	8.3 (0.19)
	Average GPAs for Students Earning More Than 10 Credits		
All students	2.70 (.65)	2.66 (.68)	2.74 (.67)
Gender			
Men	2.61 (.65)	2.61 (.68)	2.64 (.70)
Women	2.80 (.64)	2.71 (.65)	2.83 (.66)
Level of attainment			
Less than BA	2.48 (.70)	2.47 (.75)	2.43 (.72)
BA or higher	2.94 (.49)	2.88 (.51)	3.04 (.45)
Bachelor's degree major			
Business	2.78 (.49)	2.79 (.47)	2.98 (.45)
Education	2.98 (.44)	2.93 (.37)	3.16 (.39)
Engineering	2.94 (.52)	2.88 (.57)	3.02 (.46)
Physical sciences	2.94 (.49)	2.89 (.71)	3.05 (.57)
Math/Computer science	3.10 (.54)	3.02 (.50)	3.01 (.39)
Life sciences	2.98 (.48)	3.00 (.48)	3.07 (.46)
Health sciences/services	3.02 (.44)	2.90 (.43)	3.11 (.42)
Humanities	3.08 (.50)	3.04 (.45)	3.15 (.47)
Arts	3.06 (.45)	3.05 (.45)	3.14 (.47)
Social sciences	2.95 (.51)	2.85 (.56)	3.03 (.48)
Applied social sciences[3]	2.87 (.45)	2.77 (.50)	2.88 (.42)
Other	3.05 (.47)	2.86 (.54)	2.91 (.42)

[1] All undergraduate grades for all students in all institutions. Conversion from standard 4-point scale: 0-<0.7='F'; 0.7-<1.7= 'D'; 1.7-<2.7='C'; 2.7-<3.6='B'; 3.6-<=4.0='A'.
[2] Pass, Credit, (no-penalty) Withdrawal, and No-Credit Repeat grades are not included in GPA.
[3] Includes communications, public administration, criminal justice, social work, child and family services.
NOTES: (1) The universe of students whose grades are included consists of all twelfth graders in each cohort who became postsecondary participants. (2) Standard errors for grade distributions are in parentheses. (3) Standard deviations for GPAs are in parentheses.
SOURCES: National Longitudinal Study of the High School Class of 1972; High School and Beyond/Sophomore Cohort, NCES 2000-194; NELS:88/2000 Postsecondary Transcript Files, NCES 2003-402.

Table 2.2
Change in Average Undergraduate GPAs for 1972, 1982, and 1992 Twelfth Graders Who Subsequently Earned More Than 10 Credits, Measured by Effect Size

	Class of 1972 versus Class of 1982	Class of 1982 versus Class of 1992	Class of 1972 through Class of 1992
All students	–0.06	+0.12	+0.06
Men	0.00	+0.04	+0.04
Women	–0.14	+0.18	+0.04
No bachelor's earned	–0.01	–0.05	–0.07
Bachelor's earned	–0.06	+0.34	+0.21

SOURCES: National Longitudinal Study of the High School Class of 1972; High School and Beyond/Sophomore Cohort, NCES 2000-194; NELS:88/2000 Postsecondary Transcript Files, NCES 2003-402.

"Ps" or "Ws" or no-credit-repeats in the 1990s. How these would affect GPA depends on what courses (if any) were disproportionately affected and how many credits these courses typically carried (see Table 2.5).

How large were these changes in average GPA? Table 2.2 sets them out by effect size[13] across the three cohorts. What do we see? Using the late Howard Bowen's guide to interpreting standard deviation unit or effect size changes (Bowen 1977, 103), the modulations that move the galvanometer are those for bachelor's degree recipients and for women, for both of whom we mark (1) a small decline in average GPA from the 1970s to the 1980s, followed by (2) a moderate increase in average GPA from the 1980s to the 1990s.

Determined to dismiss this evidence, Zirkel (1999) denies the student as the unit of analysis—but that is what a longitudinal study is about, and the student is the one whose work is subject to judgment. Zirkel also complains that the High School and Beyond/Sophomore cohort data set of the 1980s, which showed a lower average GPA than that of the NLS-72 of the 1970s, included "a higher proportion of non-collegiate institutions." But it turns out that grades are a lot higher in cosmetology schools, for example, than they are in either community colleges or four-year colleges.[14] If these data sets were overflowing with sub-baccalaureate trade school credits, then the GPAs of Table 2.1 would be immeasurably higher. Zirkel also grieves that the High School and Beyond/Sophomore cohort "represented a wider sample of academic ability." Of course it did! That is just the

point of what happened throughout U.S. higher education in the 1980s, when the undergraduate population expanded dramatically, and that's why we include standard deviations in reporting means. It's a matter of basal statistics.

The more notable phenomena in changes of grading practices from the 1970s through the1990s is the growing proportion of withdrawals (Ws) and no-credit repeats (abbreviated as NCRs in this document), both of which are now treated as nonpenalty grades by many institutions.[15] There is an unhappy paradox here, however: what is labeled "nonpenalty" actually involves a more subtle penalty. The time one loses in such situations is time one must recoup at a later point. As Table 2.3 (using the Class of 1992) demonstrates, the volume of no-penalty Ws and NCRs has an inverse relationship to the highest degree earned. Furthermore, there is a direct relationship between the number of these grades and time-to-degree among those who earned bachelor's degrees: those with no Ws or NCRs finished in an average elapsed time of 4.13 calendar years (s.e. = 0.040), while those with seven or more such grades took an average of 6.02 calendar years (s.e. = 0.102) to complete their degrees. Other features of student history interact with the volume of Ws and NCRs, and analysts are invited to consider, for example, the distribution of such grades by institutional selectivity (Table 2.4) and secondary school course taking (Adelman 1999, 1999a). The issue of excessive nonpenalty withdrawals and no-credit repeats is more serious than their effect on grades. It is outright wastage. Think of it as 8 percent of your tuition bill and an 8 percent reduction in available class seats at most of the places that America goes to college. The higher the proportion of these grades, the greater the cost to both public subsidies for higher education and general access. In the context of our search for the "this" of grade inflation, the higher the proportion of "Ws" and "NCRs," the more traditional grades are devalued. Less than 4 percent of respondents to the American Association of Collegiate Registrars and Admissions Officers' (AACRAO) 2002 survey of grading practices indicated their institutions did *not* allow repeats, fully 55 percent indicated that a student could repeat *any* course for a better grade, and 55 percent indicated that students could repeat a course as often as they liked (AACRAO 2002). With the possible exception of the mastery learning approach to remediation in community colleges (see text that follows),

Table 2.3
Relationship of Number of No-Penalty Course Withdrawals (Ws) and No-Credit Repeats (NCRs) to Highest Degree Earned for 1992 Twelfth graders, 1992–2000

	Percent of students whose highest degree was . . .						
Number of Ws and NCRs	None	Certificate	Associate's	Bachelor's	Post-Baccalaureate	Graduate	Percent of students in category
None	22.8	10.5	8.6	36.9	11.1	10.2	32.7
	(1.27)	(0.89)	(0.77)	(1.48)	(0.73)	(0.79)	(0.85)
1–2	31.9	4.1	9.4	37.0	10.5	7.1	28.4
	(1.55)	(0.66)	(0.94)	(1.55)	(0.83)	(0.70)	(0.75)
3–6	43.1	4.3	9.7	33.0	7.0	2.8	24.4
	(2.07)	(0.99)	(1.05)	(1.71)	(0.70)	(0.47)	(0.79)
7 or more	61.0	3.1	10.6	21.5	3.5	0.3	14.6
	(2.38)	(0.93)	(1.70)	(1.74)	(0.68)	(0.17)	(0.67)

NOTES: (1) Universe consists of 1992 twelfth graders who subsequently entered post secondary education. Weighted N for highest degree = 2.09M. (2) Weighted N for bachelor's recipients for whom time-to-degree could be determined = 920k. (3) Rows for highest degree earned and columns for percent of all students may not sum to 100.0 percent because of rounding. (4) Standard errors are in parentheses.
SOURCES: NELS:88/2000 Postsecondary Transcript Files, NCES 2003-402.

repeat policy may, in fact, be more indicative of "lax" grading practices than the distribution of traditional letter grades.

For both the Class of 1982 and the Class of 1992, Table 2.4 presents the distribution of letter grades by institutional selectivity. Because students may attend institutions of varying selectivity during their undergraduate careers and also may take a course more than once, this table uses the course—not the student—as the unit of analysis. We are counting grades issued, and with all non-letter grades converted to a letter-grade schema. The reader will note the following changes in this distribution between the Class of 1982 and the Class of 1992 (the only reason for not including the Class of 1972 as well is space):

- In highly selective institutions (accounting for less than 4 percent of all grades in both cohorts), a notable increase in the proportion of "P" grades—principally at the expense of "B" grades.
- At selective institutions, an increase in the proportion of "A" and "P" grades—at the expense of "B" and "C" grades. Selective institutions increased their share of all grades from 10.5 to 16.5

Table 2.4

Distribution of Undergraduate Grades by Institutional Selectivity: All Institutions Attended by Twelfth Graders in the Class of 1982 (1982–1993) and Twelfth Graders in the Class of 1992 (1992–2000)

	Percentage of undergraduate grades that were . . .							
	A	B	C	D	F	P	WRPT[1]	Percent of all grades
Institutional selectivity								
Highly selective								
Class of 1982	30.8	42.1	15.2	2.3	2.0	6.7	1.5	2.7
	(1.61)	(1.40)	(1.12)	(0.44)	(0.31)	(0.76)	(0.25)	(0.36)
Class of 1992	31.4	33.6	14.8	2.6	1.2	14.9	1.7	3.8
	(2.07)	(1.14)	(1.62)	(0.87)	(0.33)	(1.88)	(0.17)	(0.53)
Selective								
Class of 1982	26.4	38.3	21.0	4.4	2.8	3.4	3.7	10.5
	(1.01)	(0.75)	(0.76)	(0.25)	(0.23)	(0.30)	(0.27	(0.60)
Class of 1992	30.4	33.8	16.9	3.8	2.5	8.3	4.2	16.5
	(0.85)	(0.58)	(0.59)	(0.22)	(0.20)	(0.45)	(0.29)	(0.82)
Nonselective								
Class of 1982	24.6	33.3	23.7	6.4	4.7	2.1	5.3	57.9
	(0.42)	(0.32)	(0.28)	(0.16)	(0.17)	(0.14)	(0.16	(0.90)
Class of 1992	29.2	30.8	19.0	5.0	4.1	5.6	6.4	51.9
	(0.48)	(0.29)	(0.32)	(0.12)	(0.18)	(0.18)	(0.17)	(1.02)
Open door								
Class of 1982	24.2	28.6	20.8	5.6	6.4	2.3	12.1	25.5
	(0.55)	(0.44)	(0.39)	(0.18)	(0.25)	(0.53)	(0.41)	(0.70)
Class of 1992	23.4	24.7	18.4	4.9	7.2	5.4	16.0	25.4
	(0.58)	(0.44)	(0.34)	(0.17)	(0.31)	(0.24)	(0.51)	(0.82)
Not rated								
Class of 1982	34.4	31.2	17.6	4.1	3.0	7.2	2.4	3.5
	(2.14)	(1.59)	(1.62)	(0.51)	(0.51)	(1.62)	(0.65	(0.26)
Class of 1992	36.1	33.1	14.4	3.1	2.1	6.9	4.3	2.4
	(2.24)	(1.58)	(1.42)	(0.44)	(0.43)	(0.89)	(0.91)	(0.25)

[1]Withdrawals and No-Credit Repeats combined.

NOTES: (1) All penalty grades are included under "F." (2) Rows may not sum to 100.0 percent because of rounding. (3) All undergraduate students included. (4) Standard errors are in parentheses.

SOURCES: National Center for Education Statistics: High School & Beyond/Sophomore Cohort, NCES 2000-194; NELS:88/2000 Postsecondary Transcript Files, NCES 2003-402.

percent between the Class of 1982 and the Class of 1992 as a result of both increases in enrollments in flagship state universities and changes in the selectivity ratings of some state universities from nonselective to selective.

- The proportion of Bs declined in all categories of institutions except those that were "not rated."
- Open-door institutions were by far the leaders in the proportion of withdrawals (Ws) and no-credit repeats (NCRs). Why? Because No Credit Repeats, in particular, are a staple of grading policy in remedial courses—and more remedial course work was pushed into the community colleges in the 1990s. Most community colleges require demonstration of competency in the remedial course—no matter how many times the course must be repeated—before the student can move to the "gateway" courses and on into more advanced curricula.

We have noted that Table 2.4 evidences a major shift in the distribution of letter grades in four-year colleges, regardless of selectivity level, toward the grade of P (or CR in some institutions), hence removing those grades from the calculations of GPA. The highest volume courses[16] with significant increases in the percentage of "P" grades evidence some very distinct themes, and these are illustrated in Table 2.5. All cooperative education, internship, externship, and practicum grades in four-year colleges stand out in this manner. Why custom-and-usage in grading these experiential learning courses changed so dramatically across the full range of disciplines is certainly a subject worthy of further investigation. Remedial courses delivered in four-year institutions also evidence significant increases in the percentage of "P" grades, as do a host of high-volume fractional or 1-credit courses in physical education activities and student service/orientation courses (not in table).

If there were no "P" grades, all those fractional and 1-credit courses would be included in GPAs, as would 6- or 8-credit internship or cooperative education placements. Contrary to Birnbaum's (1977) hypothesis, "P" grades do not raise grade point averages. On balance, the shift to "P" grades between the 1980s and 1990s reduces the overall proportion of As in the distribution since, in the earlier decade, over 50 percent of grades in the experiential learning internships and physical education activities were "As."

Table 2.5
Change in the Percentage of Grades Indicated by "P" (Pass, Credit) in the Undergraduate Histories of the High School Class of 1982 (through 1992) and the High School Class of 1992 (through 2000) at Four-Year Colleges, by Type of Course

	Percent of grades that were "P"	
	Class of 1982	Class of 1992
Coops and internships		
Business internship	8.5	48.5
Communications internships	6.1	36.2
Student teaching/practicums	20.3	51.9
Engineering coop	31.7	71.1
Allied health clinical externships	3.2	70.8
Psychology field work/internship	4.3	52.4
Social work practicums	5.4	14.3
Political science internships	19.4	45.4
Remedial courses (non-additive)		
Remedial reading	2.8	19.9
Remedial writing/language arts	3.8	29.5
Basic algebra	5.4	21.1

SOURCES: High School and Beyond/Sophomore Cohort, NCES 2000–194; NELS:88/2000 Postsecondary Transcript Files, NCES 2003-402.

To bring the role of institutional selectivity into the analysis of student-level performance, Table 2.6 sets forth bachelor's degree recipients' GPAs over all three grade-cohort longitudinal studies by selectivity of the institution awarding the degree. The reader will note the consistency of the relationship in both the Class of 1972 and the Class of 1982: the more selective the institution granting the bachelor's degree, the higher the student's GPA (no matter how many other institutions—some of differing selectivity—the student had attended). No similar conclusion can be reached for the Class of 1992, however. The casual observer would nonetheless say that when critics complain about all those high grades at Williams or Princeton, they are observing nothing that is particularly new.

PRIVATE AND PUBLIC COMPRESSION: PROBATION NOTICES, DEAN'S LIST, AND *CUM LAUDES*

Chan, Hao, and Suen (2005) developed a nontemporal economic model of grade inflation based on the assumption that an institution (or

Table 2.6
Undergraduate Grade Point Averages (GPAs) of 1972, 1982, and 1992 Twelfth Graders Who Earned Bachelor's Degrees, by Selectivity of the Institution Awarding the Bachelor's Degree

	Class of 1972 (1972–1984)			Class of 1982 (1982–1993)			Class of 1992 (1992–2000)		
	GPA	S.D.	s.e.	GPA	S.D.	s.e.	GPA	S.D.	s.e.
Selectivity of Institution Awarding the Bachelor's									
Highly selective	3.17	.51	.049	3.09	0.61	.070	3.15	0.48	.058
Selective	3.01	.50	.025	2.94	0.50	.028	3.07	0.44	.021
Nonselective	2.92	.45	.009	2.85	0.50	.013	3.00	0.46	.012

NOTES: (1) Weighted Ns: Class of 1972 = 733k; Class of 1982 = 855k; Class of 1992 = 921k.
SOURCES: National Longitudinal Study of the High School Class of 1972; High School and Beyond/Sophomore Cohort, NCES 2000-194; NELS:88/2000 Postsecondary Transcript Files, NCES 2003-402.

school or department within an institution) will award higher grades to render its mediocre students more competitive in the labor market, creating an "exaggerated representation of underlying fundamentals" analogous to stock analysts' recommendations or those of reviewers of audio equipment in specialized consumer magazines. Just as the ratio of "buy" to "hold" or "sell" recommendations is on the order of 9:1, so nearly all students are awarded As, and the market is fooled by the compressed signals, the authors hold.

Institutions of higher education employ other types of compressed signals than letter grades to identify both exemplary and poor academic performance. These signals are part and parcel of the landscape of judgment and unfortunately do not make any appearance in Chan, Hao, and Suen's otherwise creative analysis. Some of these signals are very public; some are not. Phi Beta Kappa keys and graduation with honors (the *cum laudes*) are obvious cases of public signals. Academic probation and academic dismissal are not likely to be publically advertised, though these notations do appear on transcripts. In both cases, we are dealing with compression of performance judgments moving grade distributions to the back burner. How do we know that they exist, let alone have any estimates of their frequency and magnitude?

In the course of building the postsecondary transcript files for the NELS:88/2000, we noticed entries on the 60,000 pieces of paper with which we were dealing that indicated academic probation, academic

dismissal, dean's list, and graduation with honors. From these, potential performance variables were constructed. Two of these variables have been used only once before, in probationary status (Adelman 2005): (1) whether the student had attained dean's list status at any time, and (2) whether the student had been placed on academic probation or been dismissed for academic reasons at any time. These are probationary variables because it is not clear what proportion of two-year and four-year institutions from which transcripts were received for the NELS:88/2000 postsecondary files will enter such information on student records.[17]

In a 2002 survey of 1,036 of its member institutions, the American Association of Collegiate Registrars and Admissions Officers (AACRAO 2002) found that 60 percent indicated "academic ineligibility to enroll" on transcripts, a phrasing that implies dismissal, not probation. In its 2005 survey of transcript practices, the AACRAO found that 70 percent of responding institutions entered academic probation and/or academic dismissal on transcripts and 86 percent indicated graduation with honors (*cum laude*, *magna cum laude*, *summa cum laude*, and their equivalents), while only 58 percent marked terms in which the student achieved dean's list status. With the exception of deans' list, these proportions are fairly robust, but all three measures will need further probing to provide guidelines for adjusting responses to mitigate any statistical biases. When Dean's list and probation/dismissal variables were tested in logistic accounts of the associate degree attainment of community college students (Adelman 2005), neither reached a level of acceptable significance. The reader should thus be cautious, and not over-interpret what one observes in Table 2.7.

However cautious we might be in assessing the data in Table 2.7, the percentages are not surprising—except, perhaps, to those who believe that higher education has no standards. If the distribution of letter grades was as outrageous at the critics claim, then we would not see one out of six postsecondary students receiving academic probation or dismissal notices. Nor would we witness *only* one out of four achieving dean's list status. As for average GPAs of these students, whether or not they earned any credentials, we find 3.29 (s.e. = 0.13) for dean's list students and 1.79 (s.e. = 0.39) for those subject to academic probation/dismissal—both of which indicate, prima facie, a high degree of reliability in institutional decision rules for recognizing

Table 2.7
Percent of 1992 Twelfth Graders Who Entered Postsecondary Institutions and Subsequently (1) Were Placed on Academic Probation at Least Once or Dismissed for Academic Reasons, and (2) Earned Dean's List Status at Any Time, by Type of Institution First Attended

	Type of first institution of attendance			
	Four-year college	Community College	Other Sub-baccalaureate institution	All
Academic probation/ dismissal	17.0 (0.97)	21.4 (1.47)	6.5 (1.34)	18.3 (0.81)
Earned dean's list status	33.9 (1.04)	17.0 (1.19)	9.5 (1.88)	25.5 (0.81)

NOTES: (1) Standard errors are in parentheses. (2) Weighted Ns: probation/dismissal = 381k; dean's list = 461k.
SOURCE: NELS:88/2000 Draft Postsecondary Transcript Files for NCES 2003-402.

superior academic achievement and flagging the lagging. To be sure, these data are not directly part of the inflation story, rather they are important patterns in the tapestry of academic performance signals. And they do remind us of the critical criterion of the reliability of grading practices within institutions, even though reliability coefficients for grades in different courses within the same department may vary considerably (Smith 1992).

As for the signal compression involved in more public recognition of achievement, consider that a Phi Beta Kappa key is awarded to a maximum of 15 percent of a graduating baccalaureate class, and those receiving the award are very likely to enter the fact on their resumes. The Phi Beta Kappa list also is very likely to be included on commencement programs. An institution may present a graduating class with an average GPA of 3.8, but the Phi Beta Kappa list always presents a public roster of the best performers. It is an accessible signal of compressed information.

Graduation with honors is another public signal of academic achievement, and 19.2 percent (s.e. = 0.82) of all bachelor's degree recipients in the NELS:88/2000 had some form of *cum laude* attached to their degrees, another fact likely to be indicated in commencement programs. Not surprisingly, the proportion of students graduating with *cum laude* or better rises directly with the selectivity of the institution awarding the degree,[18] but in the context of the grade inflation discussion the more important issue may be whether honors designations have anything

Table 2.8
Percent Distribution, by Cumulative GPA, of 1992 Twelfth Graders Who Earned Bachelor's Degrees by December 2000 with Academic Honors Versus Those Who Earned Bachelor's Degrees without Academic Honors

	Received Academic Honors	Did Not Receive Academic Honors
Cumulative GPA:		
More than 3.75	27.6 (2.23)	1.3 (0.23)
3.50–3.75	41.2 (2.35)	4.7 (0.70)
3.00–3.49	28.3 (2.16)	36.6 (1.36)
2.50–2.99	2.9 (0.79)	40.7 (1.35)
<2.50	——	16.8 (1.12)

NOTES: (1) Standard errors are in parentheses. (2) Weighted Ns: received honors = 181k; did not receive honors = 755k. (3) Columns may not add to 100.0 percent due to rounding.
SOURCE: NELS:88/2000 Draft Postsecondary Transcript Files for NCES 2003-402.

to do with GPA. Table 2.8 sets out the descriptive basics on this issue. It is another prima facie case of common sense.

Acknowledging that, according to both the 2002 and 2005 AACRAO surveys of transcript practices, a small percentage (10 percent in 2002, 14 percent in 2005) of institutions do not indicate graduation with honors on transcripts, there is still a fairly direct relationship between cumulative GPA and academic honors evident in the data reported in Table 2.8. If the average GPA of NELS bachelor's degree recipients was 3.04 with a standard deviation of 0.45 (Table 2.2), then one would expect the bulk of honors diplomas to reflect GPAs greater than 3.49—and the diplomas do.

However, in light of contentions that whole departments are known for "easy grading," and that students gravitate toward these higher grounds, one should ask after the disciplinary source of the *cum laudes* in order to judge whether some fields are overrepresented. Table 2.9 provides this information for the NELS:88/2000 cohort. A crude reading of ratios of distribution of honors diplomas to the overall distribution of majors says that academic honors signals are more noted in lower-volume fields (humanities, fine and performing arts, and physical sciences), and much less noted in two of the high-volume general fields (business and applied social sciences). As for the proportion of degree recipients within major field who emerged with academic honors, none of the comparisons involving any of the five lowest volume fields is statistically significant.

Table 2.9
Percent of 1992 Twelfth Graders Who Subsequently Earned a Bachelor's Degree with Academic Honors, by General Major Field, Compared to (1) the Distribution of All Majors and (2) the Proportion of Bachelor's Degree Recipients within Major Who Earned Academic Honors

	Distribution of Majors	Distribution of Degrees with Academic Honors	Percent of students within the Major Who Earned the Degree with Academic Honors
Social sciences	19.4 (1.01)	22.4 (1.91)	22.3 (2.00)
Business & allied[a]	17.0 (0.86)	13.2 (1.40)	15.0 (1.72)
Applied social sciences[b]	11.1 (0.76)	6.7 (0.98)	11.5 (1.77)
Education	8.7 (0.59)	10.1 (1.29)	22.4 (2.72)
Life and agricultural sciences	8.3 (0.57)	9.7 (1.59)	22.6 (3.24)
Engineering & architecture	7.9 (0.76)	7.1 (1.03)	17.2 (2.72)
Health sciences & services[c]	7.6 (0.75)	8.7 (1.63)	22.1 (3.71)
Humanities	7.0 (0.75)	8.5 (1.63)	23.3 (4.30)
Fine and performing arts[d]	5.5 (0.56)	7.4 (1.36)	25.6 (4.30)
Math and computer science	3.9 (0.56)	3.5 (0.71)	16.9 (3.82)
Physical sciences	1.6 (0.31)	2.2 (0.53)	26.4 (6.70)

[a]Includes management, accounting, marketing, administrative science, agricultural business, and so on.
[b]Includes communications, public administration, administration of justice, social work, family/community services.
[c]Includes nursing, allied health sciences, clinical health sciences, physical therapy, and so on.
[d]Also includes graphic arts/design and film studies.
Notes: (1) Columns for distribution of degrees and distribution of degrees with honors may not add to 100.0 percent due to rounding. (2) Standard errors are in parentheses. (3) Weighted N for bachelor's degrees = 920k; for bachelor's degrees awarded with honors = 177k.
Source: NELS:88/2000 Draft Postsecondary Transcript Files for NCES 2003-402.

If students migrated to majors motivated by a chance for academic honors instead of their future place in the economy or society—let alone their own interests and talents—we would witness much higher proportions of graduates with degrees in the physical sciences and fine and performing arts, for example. But the proportion of physical science graduates in the national grade-cohort longitudinal studies has declined steadily since the 1970s, while the proportion of fine and performing arts graduates has remained flat (Adelman 2004, table 5.1, p. 61). Future research certainly has some untangling tasks involving changes in the enrollment mix, average GPAs by field (see Table 2.2 in this chapter), and more systematic examination of the

compressed signaling of honors, particularly in fields that have experienced volatile changes in share of traditional age bachelor's degree recipients over the past thirty years: business, education, and engineering.

ENROLLMENT MIX AND THE PROXY MEASURES

Enrollment mix, in fact, is one of the key influences on the distribution of letter grades, and a few of those who study grading trends in relation to the production function of academic departments (e.g., Sabot and Wakeman-Linn 1991) have the sense to ask whether all those As were given in recreation or chemistry. The polemics, trade press, and general press articles and op-eds do not reflect such commonsense questions. The transcript evidence shows that grade distributions vary considerably from course to course, a finding in keeping with other research that examines grades at more discrete levels than schools or departments (e.g., Prather, Smith, and Kodras 1979). The course category with the highest percentage of A grades earned by the 1992–2000 longitudinal studies cohort is Music Performance (keyboard, vocal, woodwind, brass, percussion, opera, rock, folk, jazz). These are normally 1-credit courses. Consider the effect on the presentation of letter grade distributions inherent in the following comparison from the NELS: 88/2000 transcript files:

	Modal Credit	Cases	Percent As
Music performance	1	4725	64.2 (s.e. = 1.99)
U.S. government	3	3525	15.5 (s.e. = 0.89)

If, by enrollment mix, music is disproportionately weighted among majors at university X, then it would not be surprising to find the distribution of grades—though not necessarily average GPA—tilted toward the high end of the spectrum.

Enrollment mix, a function of the intersections of student choice, institutional requirements, course scheduling, departmental prerequisites, departmental reputation, and departmental habits in grading, also has multiplier effects. If a higher percentage of students major in psychology, the then grading habits of psychology departments will dominate grade distributions. And since 77 percent of bachelor's degrees in psychology during the period 2001–2002 were awarded to women (Snyder 2004, table 293, p. 353) the odds of an upward tilt in the

distribution are enhanced. More of X means less of Y in enrollment mix, so when a lower percentage of students are majoring in computer science, the grading habits of computer science departments will have less impact.

In general, it is held, large general education distribution courses evidence lower grade profiles, while low enrollment courses with small classes will have higher grade profiles (Dickson 1984). Table 2.10 takes eight high-enrollment-volume courses from the Class of 1992 files to illustrate the former phenomenon. Ideally, instead of displaying the entire sweep of a cohort's course taking attainment over time, one should present year-by-year distributions of grades in these eight illustrative courses. What would happen in a grade-cohort study, though, is that the volume of course-taking would drop off, particularly for field entry-level courses such as U.S. Government and Introductory Accounting. For that reason, the tracking of grade distributions in individual courses is best left to institutional-level analysis. Nonetheless, the data of Table 2.10 suggest that distributions of grades tilting toward honors will more likely be found in second- and third-level courses with prerequisites, for example, Organic Chemistry, Ethics, and Technical Writing.

Notice that nearly two out of every three grades in ethics are honors grades (A or B), compared to 45 percent in Introductory Accounting, Calculus, and U.S. Government—courses in which the withdrawal, repeat, and failure rates were comparatively high. The courses represented in Table 2.10 were taken by the NELS:88/2000 postsecondary students in a minimum of 438 institutions (for Organic Chemistry) to a high of 1,105 institutions (for Introductory Accounting). Approximately 1 million grades were recorded for introductory and intermediate-level Spanish courses, compared to 231,000 for Microbiology. These are very robust numbers that open up lines of research inquiry for those who wish to add to the analysis such institutional variables as sector, Carnegie type, and selectivity.

It is at the course level, in fact, that the field of research not only on grading practices, but more broadly on assessment practices, opens up. Table 2.10 picked high-enrollment-volume courses as its subject and deliberately selected from different sectors of the disciplinary spectrum. Researchers who wish to investigate the enduring *effects* of grades might start with students' stated intentions of major, identify a set of "gateway" courses in that field, and model the relationship

Table 2.10
Letter Grade Distribution in Selected High-Enrollment-Volume Undergraduate Courses Taken by 1992 Twelfth Graders: 1992–2000

	Introductory Accounting	Introductory And Intermediate Spanish	Technical Writing	Micro-biology	Calculus	Ethics	Organic Chemistry	U.S. Gov't
A	18.1	27.3	34.0	20.6	19.1	31.8	21.8	15.5
	(0.99)	(1.44)	(1.85)	(2.03)	(1.16)	(2.37)	(1.48)	(0.89)
B	27.6	29.5	39.1	29.7	25.7	34.5	30.8	29.5
	(1.12)	(1.07)	(1.89)	(2.64)	(1.25)	(1.99)	(1.49)	(1.14)
C	22.4	18.4	13.4	31.0	25.8	17.4	25.5	27.3
	(1.16)	(1.03)	(1.35)	(4.34)	(1.28)	(1.40)	(1.29)	(1.17)
D	7.6	5.4	2.3	4.7	7.8	4.1	5.5	8.9
	(0.62)	(0.52)	(0.58)	(0.92)	(0.94)	(0.79)	(1.29)	(0.76)
F	7.0	4.8	3.5	2.7	5.5	2.2	3.6	6.1
	(0.58)	(0.65)	(0.83)	(0.59)	(0.65)	(0.42)	(0.57)	(0.83)
Pass/credit	1.1	4.7	1.6	3.9	3.7	3.1	2.2	0.9
	(0.30)	(0.62)	(0.39)	(0.88)	(0.98)	(1.02)	(0.48)	(0.21)
Withdraw	11.1	8.1	5.6	5.5	7.4	6.1	7.1	9.2
	(0.80)	(0.70)	(0.95)	(1.05)	(0.69)	(1.17)	(0.84)	(0.90)
Repeat	4.5	1.8	0.5	1.9	6.9	0.8	8.7	2.6
	(0.45)	(0.39)	(0.18)	(0.84)	(0.70)	(0.51)	(0.63)	(0.39)
Number of Institutions	1105	959	601	513	674	699	438	1002
Weighted Cases	849k	961k	308k	231k	691k	338k	503k	827k

Notes: (1) Columns may not add to 100.0 percent because of rounding. (2) The universe and weights are those for all known postsecondary participants. (3) The threshold for "high enrollment volume" was set to 200,000 weighted cases of course-taking over the period 1992-2000. (4) Standard errors are in parentheses.
Source: NELS:88/2000 Postsecondary Transcript Files, NCES 2003-402.

between grades in the gateways to subsequent attainment. The effect of assessment practices such as oral performance in Intermediate Spanish, online simulations in Introductory Accounting, and instrumentation setup in Microbiology, each with its own set of performance criteria may prove even more worthy of exploration in terms of their relationship to grades than the most sophisticated modeling of the proxy indicators of performance themselves.

THERE IS STILL NO ANSWER, SO WHAT ARE THE ALTERNATIVES?

We have surfed through pages of data, with all their qualifications. I trust the reader recognizes that the landscape of undergraduate grades has more dimensions than something called "inflation." And while we have encountered a few challenging hypotheses and explorations in the literature, none of them tells us what the "this" of "inflation" is. Neither the literature nor the national time series data proves that U.S. higher education is paying a higher price for a lower quality product, or a higher price for a stable quality product, or even a higher price for a product that has risen in quality at a lower rate.

From this author's perspective, the alternative is to set the elusive topic of "inflation" aside and see if we can induce college faculty to discover and master criterion-referenced assessment that truly distinguishes levels of performance, thus restoring grades as information systems that speak simultaneously to the student and to the world outside the academy. The task envisions intellectual sweat and physical time, but it certainly will help those charged with both instruction and judgment determine what grades are and what we want them to do, a point eloquently made two decades ago by Milton, Pollio, and Eison (1986). I suspect that the root of contemporary little ease about grades and grading is that we do not know why we are paying the reported prices. Maybe it is time we found out.

NOTES

The majority of the data and core analyses in this chapter was originally published in C. Adelman. *Principal Indicators of Student Academic Histories in Postsecondary Education, 1972–2000.* (Washington, DC: U.S. Department of Education, 2004), 77–86.

1. The National Longitudinal Study of the High School Class of 1972 began with twelfth graders in 1972 and collected their college transcripts in 1984. The High School and Beyond/Sophomore Cohort started with tenth graders in 1982 and collected their college transcripts in 1993. The National Education Longitudinal Study of 1988 (NELS:88/2000) began with eighth graders in 1988 and collected their college transcripts in the fall of 2000. The sampling design for all these studies was the same. The raw Ns for the initial sample were in the range of 22,500 to 30,000; the weighted Ns, from 2.9 to 3.8 million.

2. In the High School and Beyond/Sophomore Cohort Longitudinal Study, for example, where both student self-reports of GPA and transcript evidence are available, students overestimated their high school GPA by .22, with measurable variation by race/ethnicity. See Fetters, Stowe, and Owings (1984) and Adelman (1999a). The use of self-reported student grades is thus not highly recommended.

3. No, they are not representative of all four-year college students or all community college students, though obviously far more representative than the self-selected fragments one finds at www.gradeinflation.com. The sampling design seeks a nationally representative body of tenth graders, for example, and then follows all of them, no matter what type of postsecondary institutions they attend (if any).

4. The National Center for Education Statistics will provide instructions and forms for licenses to use the restricted files. The CDs that should be specified for the postsecondary transcripts are #2003-402and Supplement (for the NELS:88/2000), #2000-194 (for the High School and Beyond/ Sophomore Cohort), and an unnumbered CD issued in 1994 for the National Longitudinal Study of the High School Class of 1972. To create analysis files from the data sets, the analyst should be able to program in SAS, SPSS, or STATA.

5. Rosovsky and Hartley (2002), *Evaluation and the Academy*, for example, offer 35 citations concerning "grade inflation," of which only eight were published in academic journals. The others include five newspaper stories, ten opinion pieces from the newspapers and trade press, six narratives in the trade press, and three pieces of fugitive literature for which one can find abstracts but no text.

6. The authors found the percentage of bachelor's degree recipients graduating with academic honors in thirty-three Texas colleges rising from 21 percent in 1959–60 to 33 percent in 1974–75. The NELS:88/2000 postsecondary transcript files show 19 percent (s.e. = 0.80) of a traditional-age cohort who earned bachelor's degrees between 1996 and 2000 as graduating with honors (see *www.aacrao.org/pro_dvelopments/surveys/transcript05.htm*).

7. Some grades were presented on a percentile scale, some on a 4-point scale, some with standard alphabetic symbols, some with nonstandard alphabetic symbols, and a few by narrative. In one noted case, virtually every course carried three component numerical grades; in another, three component narratives.

8. First calendar year, cumulative through two calendar years from the date of entry, and final undergraduate, whether or not a degree was earned. An additional variable was created to describe the trend (rising, flat, falling) in the student's GPA across those three reference points. The data sets are flexible enough so that year-by-year GPAs can be calculated and taken into consideration in declining risk sets commonly used in event history analyses.

9. Wegman (1987) "An Economic Analysis," contends that this process is analogous to indexing the changing cost of baskets of identical goods and services to determine the core rate of inflation in the economy writ large. His

general grade inflation index is built from total grade points awarded in a given basket of courses over a "base" period (roughly eight years) compared to total grade points in the same courses in a second period of similar length, and both method and results look reasonable. But Wegman's grade index does not get beyond the proxies of price to course material, types of assessments, and performance criteria that would be critical in appraising change in the judgment of human intellectual performance.

10. All data cited in this paragraph were generated from the National Center for Education Statistics' public release (Data Analysis System) CD for the Beginning Postsecondary Students Longitudinal Study of 1995/96-2001.

11. Lacher and Wagner (1987) "MCAT Scores to Grade Point Ratio," provide another case of post-hoc matching using subtest scores from the Medical College Admissions Test and average GPA for each institution represented by testtakers to create a comparative *institutional* "grade inflation index" as a context for interpretation of individual MCAT scores.

12. Anyone who picks a date earlier than 1972 for the reference point (1) does not have representative data (we did not collect it), (2) willingly brings the Vietnam War period distortions into the picture, and, most importantly, (3) ignores the fact that 1972 was a watershed year in which the core rules of access to higher education in the United States were altered, launching a noted period of massification. In the three years prior to 1972, total undergraduate enrollments rose 11 percent; in the three years after 1972, they rose 22 percent (Snyder, 2004, table 189, p. 238). That contrast should provide adequate hints that measures of just about anything in higher education using a baseline prior to 1972 refer to a different geological era.

13. The effect size is a simple calculation of the pooled standard deviation divided by the difference in means. It produces the same results as does the calculation for standard deviation units.

14. In the NELS:88/2000 cohort, for example, mean GPAs for all students who earned more than 10 credits were 3.10 (s.e. = 0.45) for their work in sub-baccalaureate trade schools, 2.64 (s.e. = 0.27) for their work in community colleges, and 2.75 (s.e. = 0.14) for their work in four-year colleges. The master program (written in SAS by the author) for these constructions is provided on the restricted CD (NCES 2003-402) and allows for the calculation of undergraduate GPAs at other institutional types configured by the analyst.

15. Both Birnbaum (1977), "Factors Related to University Grade Inflation," and Kolevzon (1981), "Grade Inflation in Higher Education," mention no-penalty withdrawals in their analyses, acknowledging them as part of institutional grading policy, but obviously did not have access to the time series data that the three national longitudinal studies provide so they could not see this change. However, McSpirit and Jones (1999), "Grade Inflation Rates," make note in one institution, covering the years 1983 to 1996, of the same pattern we observe nationally.

16. A threshold of 4,000 weighted cases of undergraduate "P" grades in the transcript files of the NELS:88/2000 was set for a course to be admitted for consideration in Table 2.5.

17. Received transcripts from 1,197 of the 2,557 institutions (46.8 percent) evidenced one or more of these entries, but with a very uneven distribution by institutional selectivity. It could be that the students in the NELS:88/2000 sample who attended institution X were all quite average, were never placed on academic probation and never earned dean's list status. Hence, we would never know from the received transcripts whether student records at institution X carry pejorative and/or honorific performance entries. AACRAO's continuing surveys of registrars should prove helpful in providing a complete account.

18. For students earning bachelor's degrees from highly selective institutions, 27.4 percent (s.e. = 3.97) received the degree with honors, compared with 20.8 percent (s.e. = 1.81) from selective institutions, and 17.8 percent (s.e. = 0.99) from non-selective schools.

REFERENCES

Adelman, C. 1999a. *Answers in the Tool Box: Academic Intensity, Attendance Patterns, and Bachelor's Degree Attainment.* Washington, DC: U.S. Department of Education.

Adelman, C. 1999b. *The New College Course Map and Transcript Files: Changes in Course-Taking and Achievement, 1972–1993.* 2nd Ed. Washington, DC: U.S. Department of Education.

Adelman, C. 2004. *Principal Indicators of Student Academic Histories in Postsecondary Education, 1972–2000*. Washington, DC: U.S. Department of Education.

Adelman, C. 2005. *Moving Into Town—and Moving On: The Community College in the Lives of Traditional-age Students.* Washington, DC: U.S. Department of Education.

American Association of Collegiate Registrars and Admissions Officers. (AACRAO). 2002. *Academic Transcripts and Records: Survey of Current Practices.* Washington, DC: Author.

Bearden, J., R. N. Wolfe, and J. W. Grosch. 1992. "Correcting for Grade Inflation in Studies of Human Performance." *Perceptual and Motor Skills* 74: 745–46.

Beck, B. 1999. "Trends in Undergraduate Grades." Paper presented to the Fall Conference of the Association for Institutional Research in the Upper Midwest, St. Paul, Minnesota.

Birnbaum, R. 1977. "Factors Related to University Grade Inflation." *Journal of Higher Education* 48:5: 519–39.

Bowen, H. R. 1977. *Investment in Learning.* San Francisco: Jossey-Bass.

Bromley, D. G., Crow, M. L. and Gibson, M.S. (1978).Grade Inflation: Trends, Causes, and Implications. Phi Delta Kappa. 59, 694–97.

Chan, W., L. Hao, and W. Suen. 2005. "A Signaling Theory of Grade Inflation." Toronto, Ontario: University of Toronto Working Paper.

Dickson, V. A. 1984. "An Economic Model of Faculty Grading Practices." *Journal of Economic Education* 15:3:197–204.

Farkas, G., and L. Hotchkiss. 1989. "Incentives and Disincentives for Subject Matter Difficulty and Student Effort: Course Grade Determinants across the Stratification System." *Economics of Education Review* 8:2:121–32.

Fetters, W. B., Stowe, P. S., and Owings, J. A. 1984. Quality of Responses of High School Students to Questionnaire Items. Washington, DC: National Center for Education Statistics.

Freeman, D. G. 1999. "Grade Divergence as a Market Outcome." *Journal of Economic Education* 30:4:344–51.

Kolevzon, M. S. 1981. "Grade Inflation in Higher Education: A Comparative Study." *Research in Higher Education* 15:3:195-212.

Kuh, G. D. and Hu, S. 1999. "Unraveling the Increase in College Grades for the mid-1980s to the mid-1990s." Educational Evaluation and Policy Analysis, vol. 21, no. 3, pp. 297–320.

Lacher, D. A., and S. M. Wagner. 1987. "MCAT Scores to Grade Point Ratio: An Index of College Grade Inflation." *College and University* 62:3:201–206.

Levine, A. and Cureton, J. S. 1998. When hope and Fear Collide: a Portrait of Today's College Student. San Francisco: Jossey-Bass.

McKenzie, R. B. and R. J. Staaf. 1974. *An Economic Theory of Learning.* Blacksburg, VA: University Publications.

McSpirit, S., and K. E. Jones. 1999. "Grade Inflation Rates among Different Ability Students, Controlling for Other Factors." *Education Policy Analysis Archives* 7:30:15. *http://www.epaa.asu.edu/epaa/v7n30.html.* accessed.

Michaels, J. W., and T. D. Miethe. 1989. "Academic Effort and College Grades." *Social Forces* 68:1:309–19.

Millman, J., S. P. Slovacek, E. Kulick, and K. J. Mitchell. 1983. "Does Grade Inflation Affect the Reliability of Grades?" *Research in Higher Education* 19:4:423–29.

Milton, O., H. R. Pollio, and J. A. Eison. 1986. *Making Sense of College Grades.* San Francisco: Jossey-Bass.

Olsen, D. R. 1997. "Grade Inflation: Reality or Myth?" Paper presented at the Annual Forum of the Association for Institutional Research, Orlando, Florida.

Oltman, P. K. 1982. *Content Representativeness of the GRE Advanced Tests in Chemistry, Computer Science, and Education.* Princeton, NJ: Educational Testing Service.

Prather, J. E., G. Smith, and J. E. Kodras. 1979. "A Longitudinal Study of Grades in 144 Undergraduate Courses." *Research in Higher Education* 10: 1:11–24.

Rosovsky, H., and M. Hartley. 2002. *Evaluation and the Academy: Are We Doing the Right Thing?* Cambridge, MA: American Academy of Arts and Sciences.

Sabot, R., and J. Wakeman-Linn. 1991. "Grade Inflation and Course Choice." *Journal of Economic Perspectives* 5:1:159–70.

Smith, D. L. 1992. "Validity of Faculty Judgments of Student Performance." *Journal of Higher Education* 63:3:329–40.

Snyder, T. 2004. *Digest of Education Statistics, 2003*. Washington, DC: National Center for Education Statistics.

Stone, J. E. 1995. "Inflated Grades, Inflated Enrollment, and Inflated Budgets: an Analysis and Call for Review at the State Level." Education Policy Analysis Archives [that should be in italics], vol. 3, no. 1, June 26, 1995.

Summerville, R. M., and D. R. Ridley. 1990. "Grade Inflation: The Case of Urban Colleges and Universities." *College Teaching* 38:1:33–38.

Wegman, J. R. 1987. "An Economic Analysis of Grade Inflation Using Indexing." *College and University* 62:2:137–46.

Zirkel, P. 1999. "Grade Inflation: A Leadership Opportunity for Schools of Education?" *Teachers College Record* 101:247–60.

CHAPTER THREE

Understanding Grade Inflation

RICHARD KAMBER

FIVE QUESTIONS

American educators today are deeply divided over the extent to which grade inflation poses a problem for higher education and what, if anything, should be done about it. Some, like Henry Rosovsky, former dean of the Faculty Arts and Sciences at Harvard, and Matthew Hartley, currently assistant professor at the University of Pennsylvania's Graduate School of Education, see it as a self-sustaining hazard that "weakens the very meaning of evaluation."[1] Others believe it is harmless as long as professors maintain high classroom standards and effectively assess student work.[2] A few, like Clifford Adelman and Alfie Kohn, have questioned the existence of grade inflation as a nationwide phenomenon. Adelman, a senior research analyst at the U.S. Department of Education, has spoken of "the folklore of grade inflation" and has argued that between 1972 and 1993 "the mean GPA for students earning more than 10 credits went from 2.80 to 2.66."[3] In 1995, he claimed that "at most schools there is no grade inflation."[4] In 2001, he argued: "The case for grade inflation in U.S. higher education that is played out in the media is not based on anything you or I would regard as inflation."[5] Not until 2004 did he concede a rise in the proportion of A grades between 1982 and 1992.[6] Kohn, a popular writer on the education of children, has labeled grade inflation "a dangerous myth" and has challenged "the simple claim that grades have been rising." He has said: "Depending on the time period we are talking about, that claim may well be false."[7]

How is it possible for scholars to be so far apart on the problem of grade inflation? The answer I believe lies in the inherent complexity of the problem, the misconceptions that have grown up around it, and the self-deception of educators who prefer denial to cure. The aim of this chapter is to provide an overview of the problem that does justice to its complexity and points toward realistic remedies. In pursuit of this aim I borrow a model from the field of public health. When epidemiologists confront a public health problem, be it AIDS, measles, crack cocaine, binge drinking, or obesity, the questions they ask typically include the following:

- What are the symptoms and defining characteristics of grade inflation?
- What is the etiology (origin and causes) of grade inflation?
- Why is grade inflation harmful?
- Whom does grade inflation afflict?
- What can be done to contain, ameliorate, or eradicate it?

By examining grade inflation as though it were a public health problem and offering fresh answers to the first four questions I try to dispel the confusion that has prevented educators from reaching consensus on what, if anything, to do about this problem. In Chapter 9, "Combating Grade Inflation: Obstacles and Opportunities," I suggest specific remedies for ameliorating and reducing grade inflation.

WHAT ARE THE SYMPTOMS AND DEFINING CHARACTERISTICS OF GRADE INFLATION?

The symptoms of grade inflation are familiar: an upward shift in the grade point average of students over an extended period of time, typically marked by a notable increase in As and Bs and a notable decrease in grades from C on down. However, there is disagreement among scholars as to whether grade inflation can be defined simply as "an upward shift in the grade point average of students over an extended period of time"[8] or whether one needs to add "without a corresponding increase in student achievement."[9] The point of adding "without a corresponding increase in student achievement" is to suggest that what is objectionable about rising grades is not merely that they are rising but that they are rising without justification.

My dissent is more radical. I believe that in order to expose what is objectionable about rising grades, grade inflation is best defined as a "reduction in the capacity of grades to provide true and useful information about student performance as a result of upward shifts in grading patterns." In other words, I believe the root problem is loss of information to students rather than lack of justification by teachers. A quick way to understand my argument is to note that a *downward* shift in grades could produce a comparable loss of information. If an appreciable number of colleges over the past forty years had adopted the practice of giving undergraduates mostly Ds and Fs rather than mostly As and Bs, then the reduction in reliable and useful information about student performance would have been much the same. A good name for *any* grading practice that diminishes the capacity of grades to provide students with useful information is "grade conflation."[10]

My definition of grade inflation deliberately omits a reference to grade point average (GPA). Although a rise in grade point average is a necessary concomitant of net upward shifts in grading, I think an increase in GPA should be treated as a measure of grade inflation rather than as part of its definition. One also can measure rising grades by tracking the distribution of As, Bs, and so on. As for "a corresponding increase in achievement," I believe it possible to have grade inflation even when there is a corresponding increase in student achievement. If this belief seems paradoxical, then the reason I submit it is not my muddled grasp of the problem but the iron grip that the misleading metaphor "grade inflation" has come to have on common understanding. I shall say more in defense of my definition later in this chapter, but first I must pry up the ferrous fingers of the metaphor.

While the problem we now call "grade inflation" began in the early 1960s, the metaphor "grade inflation" did not become the standard way of referring to it until the mid-1970s. There was nothing inevitable about this. A 1972 article by *New York Times* reporter Iver Peterson credited "the phrase" to Harvard sociologist David Riesman and helped bring it to the attention of a national audience.[11] Yet other articles written about this time named the problem "lax grading,"[12] "too many As," or "grade glut,[13] rather than "grade inflation." Probably "grade inflation" gained ascendancy because of the double-digit price inflation in the 1970s.[14] I doubt that anyone in the mid-1970s foresaw the long-term confusion this economic metaphor would engender,

though no one was better qualified to do so than Riesman. Two decades earlier, he had written in *The Lonely Crowd*, "Words not only affect us temporarily; they change us, they socialize, or unsocialize us."[15]

Today, "grade inflation" is so entrenched in American usage that we lack an alternative name. This lack would not matter much if "grade inflation" were a dead metaphor like "eye of the storm" or "arm of the chair," but "grade inflation" remains a live and systematically misleading metaphor. It evokes associations that cause its users to think about grades in ways that distort their understanding of the problem. Medical science has long recognized that popular names such as "the king's evil" for scrofula (a form of tuberculosis) or "the kissing disease" for mononucleosis can be misleading and has adopted technical names to replace them. The same correction has not been made with "grade inflation." Let us, therefore, arm ourselves against confusion by noting some of the ways in which grade inflation and price inflation are dissimilar.[16]

First, grades belong to systems that are sealed at both ends. You can charge as much as you like for a product, but you cannot award more than your system's highest grade—usually an A.[17] When an A is awarded for what was previously B-level work, the system loses its capacity to recognize the superiority of what had been A-level work. To avoid this unhappy consequence, some faculty bravely hold the line on As, while being more generous with other passing grades. But this strategy is self-defeating. When Bs, for example, are awarded for what was previously C-level work, then the only way to differentiate what was previously B-level work is to award As to that work—which then deprives the system of its capacity to recognize A-level work. And so on down the line: the awarding of *any* passing grade in the system to student work that was previously graded at a lower level creates a "recognition deficiency" that can be remedied only by passing on the problem to the grade above it. Like a bubble in a sealed envelope, this "recognition deficiency" can be left bulging at a lower level, squeezed to the top, or spread out in some fashion across the system—but the bubble remains.

Second, buyers may bargain for prices, but it is sellers who set them. We tell a housepainter what we want him to do; he tells us how much it will cost. When we tell a student to write a term paper, we do not expect, nor should we tolerate, the demand of a set grade in return. As buyers of a college education, students are entitled to

demand that we grade their work and do so promptly and fairly, but that is in their capacity as buyers of an education, not as sellers of assigned work. We do not buy work from students, and the particular grades that we award are tokens of recognition, not means of exchange.[18]

Third, as economists Paul Samuelson and William Nordhaus have pointed out: "There is no effect on real output, efficiency or income distribution of an inflation that is both balanced and anticipated."[19] Grades are different: the ills of grade inflation cannot be offset by "keeping pace" with it. A teacher who starts giving higher grades to keep pace with her colleagues may make her students happier in the short run, but she adds to the cumulative harm of grade inflation.

Fourth, price inflation is remedied by bringing it down to a low, steady, and predictable rate. Rolling back prices is seldom necessary or even desirable. If the United States were suddenly faced with a new surge of inflation, it would be folly to think of driving prices down to the levels that prevailed in 1960 rather than slowing the rate of increase and rectifying imbalances in wages, pensions, taxes, and so on. Not so with grade inflation: the damage it does is cumulative and can be combated only by restoring meaningful distribution of grades. In 1960, grades provided fairly reliable information about a student's academic performance relative to fellow students, regardless of the college from which they came. An A or A– in a course placed a student's work in the top 10–20 percent; a C or C+ meant that a student's work was about average. Today, there is far less commonality. At some colleges (but not others) an A or A– places a student's work in the top 45 percent, while a C or C+ means little more than "showed up and completed most assignments." Even if grade inflation were now at a standstill, grading practices in American higher education would need to be reformed.

In short, the name "grade inflation" is about as helpful for understanding the problem to which it refers as the name "king's evil" is for scrofula, or "the kissing disease" is for mononucleosis. The key to correct diagnosis is the recognition that grading systems are first and foremost systems for providing students with information rather than systems of exchange or reward. Grades are informative to the extent that those who see them (students, parents, employers, graduate schools, etc.) know what they signify about student performance and to the extent that what they signify is worth knowing. But grades lose their capacity to provide information when their distribution is at odds

with the conventions that give them meaning, or when grades are used to signify very different things about student performance. A grade of B, for example, can no longer serve its historic function of signifying above-average work when it has become a below-average grade. A grade of A ceases to have clear significance when it is used to indicate that a student has done better than her classmates *or* improved a great deal *or* tried very hard *or* had to finish her work while coping with a family crisis. The root problem with this disjunctive A is not that it is too lenient but that it is insufficiently informative.

WHAT IS THE ETIOLOGY (ORIGIN AND CAUSES) OF GRADE INFLATION?

Grade inflation as a local phenomenon has many origins. Wherever teachers have had discretion in the awarding of grades, some have been easy graders. Even in the heyday of normal curve grading, students in search of high marks without hard work found "gut" or "cake" courses to help meet their needs. But grade inflation in epidemic proportions is another matter. The most common view is that grade inflation became epidemic in the late 1960s as a response to student unrest and growing disaffection with establishment values. My uncommon view is that this was the second rather than the first epidemic. I believe there is sufficient evidence to infer that something akin to grade inflation reached epidemic proportions in the early nineteenth century and lasted for over fifty years.

I say "something akin to grade inflation" because there was no standard measure of GPA during this period, and few grading records have survived. But grades in one form or another were a consistent feature of American higher education from its very beginning, and students needed to earn passing grades in order to graduate. Students at Harvard College in 1646 had to prove that they were able "to read y^{e} originall of y^{e} old and new testaments into y^{e} Latin tongue, and to Resolve them Logistically."[20] Whatever the limitations of America's colonial colleges (and there were many), they seem to have maintained sufficient standards to produce well-read and articulate clergy and local leaders. Thirty of the fifty-five delegates to the Constitutional Convention were college graduates.

In the early decades of the nineteenth-century American higher education grew so rapidly that it outstripped its capacity to maintain

even the modest standards of the colonial period. As historian of education Frederick Rudolph observes: "The American people went into the American Revolution with nine colleges. They went into the Civil War with approximately 250."[21] In addition, "Perhaps as many as seven hundred colleges tried and failed before the Civil War."[22] Rudolph explains these astonishing numbers by the entrepreneurial spirit of their founders and the small size of the institutions they created. "College-founding in the nineteenth century was undertaken in the same spirit as canal-building, cotton-ginning, and gold-mining."[23] In 1846, New York City's two colleges (Columbia and the University of the City of New York—later NYU) enrolled a total of 247 students; Lafayette's student body in 1848 was smaller than its board of trustees; Denison graduated just sixty-five students in its first *twenty* years of operation; Harvard graduated its first class of 100 in 1860—and these were the successful institutions. Clearly, few colleges could afford the luxury of failing students. Student letters suggest the academic standards of the time. In 1822, Mark Hopkins, then a student at Williams, wrote to his brother: "The members of the Senior class have all passed their examinations successfully as is usual on such occasions. They say that the Reverend and Honorable examiners . . . nodded and slept half the time."[24] Rudolph concludes that "nowhere were students being dismissed for academic reasons and nowhere were really challenging intellectual demands being placed upon them."[25] As late as 1894, Harvard's Committee on the Raising the Standard reported "that in the present practice Grades *A* and *B* are sometimes given too readily—Grade *A* for work of no very high merit, and Grade *B* for work not far above mediocrity."[26]

Elsewhere in this volume, Mary Biggs's "Fissures in the Foundation" (Chapter 7) traces the history of grading in American higher education from its beginning to the present. She establishes that normal curve grading (with some skewing toward the upper end of the range, awarding more As and Bs than Ds and Fs) became the accepted standard early in the twentieth century and remained the standard until the mid-1960's. She also points to a lag of several years between the beginning of the 1960s grade inflation epidemic and its recognition by the educational establishment or the popular press. I believe this lag was due in part to the difficulty of finding adequate data. In the absence of national surveys of grading practices by federal agencies or national organizations, investigators had to rely on data gathered

serendipitously by various consortia or what they themselves could pry loose from registrars. At a time when many institutions were still working with handwritten records, obtaining longitudinal reports on grading trends was no easy matter. The articles that surfaced in *Newsweek*, *Time*, and the *Wall Street Journal* in the early 1970s relied on statistics from small numbers of institutions.

The first scholarly paper to make a statistically credible case for a national epidemic of grade inflation was Arvo Juola's 1976 article, "Grade Inflation in Higher Education: What Can or Should We Do?"[27] Using a stratified sample of 485 institutions with usable returns from 134 (28%), Juola traced an increase of GPA from 2.4 in 1960 to 2.8 in 1973, with two-thirds of this rise occurring from 1968 to 1973. What amazed Juola was that the "same trend and nearly the same magnitude of change was evident" in institutions of every kind, size, and geographic location.[28] When he compared this trend with annual changes in GPA at Michigan State University since 1941, he found that "the increments in the 1960–65 period, and, perhaps, even the somewhat larger gains in the 1965–67 period, show normal fluctuations in grade trend levels"[29] (fluctuations in the range from 2.3 to 2.5 with "a low ebb in grading levels in the post-Soviet Sputnik era"[30]) "until 1968 when it jumped to 2.56 and then continued to near the 2.8 mark in the early '70s."[31] In a 1979 study, "Grade Inflation—1979. Is It Over?," Juola reports that the average GPA (at institutions granting graduate degrees) peaked in 1974, declined by .043 from 1974 to 1977, and then rose by .001 from 1977 to 1978, for a net loss of .042.[32] He reports an average GPA of 2.72 for "all colleges" in 1978.[33]

Various causes and clusters of causes have been cited for the surge of grades from 1968 to 1974. One cluster is political and ideological. It is widely surmised and probably true that a significant percentage of instructors (especially younger instructors) softened their grading practices to show solidarity with students whose lives were being shaken by forces from outside the academy: the escalating war in Southeast Asia, the draft, the radicalization of the civil rights movement, the assassinations of Martin Luther King and Robert F. Kennedy, the shooting of students at Kent State and Jackson State, the counterculture, the women's movement, and the sexual revolution.[34] This was a moment of deep dissent in American society. Students began to demand "relevance" in the classroom, participation in college and

university governance, and greater freedom in the conduct of their personal lives. Rules and standards of every kind were subject to student criticism and negotiation.

A second cluster of causes is market-related, demographic, and fiscal. Institutional overexpansion, declining applicant pools, and cuts in budgetary support for higher education in the early 1970s prompted trustees, administrators, and faculty to be more permissive in recruiting and retaining students.

A third cluster includes internal policy changes that made it easier for students to avoid being penalized by low grades. Many schools from 1968 to 1974 adopted pass/fail options, the repeating of courses with new grades substituted for old, the expansion of withdrawal options, the inclusion of transfer grades in cumulative GPAs and honors calculations, and the exclusion of Fs from these calculations.[35] A related change was the institution of student evaluations of faculty. Although many years would pass before the influence of student evaluations on grades was adequately documented, the influence itself was probably immediate.

In addition to these potent clusters, other—more questionable—causes have been adduced for America's second grade inflation epidemic. David Riesman's examination of grade inflation in his influential overview of American higher education for the Carnegie Council (*On Higher Education*, 1980) emphasized:

> [t]he pressure to give course credit for remedial [and pre-college work], first demanded by and on behalf of unprepared blacks and then by some other non-whites similarly handicapped by family background and the inadequacies of previous schooling."[36]

Riesman hypothesizes that this pressure contributed both directly and indirectly to grade inflation.

> Indeed, the presence of visibly ill-prepared non-whites (as against the invisible equally unprepared whites scattered in almost any classroom) led many guilty or intimidated white instructors to bestow passing grades or even honor grades on inadequate work."[37]

Similar claims have been made from time to time. Harvey Mansfield,

a political scientist at Harvard, drew charges of racism by reaffirming this hypothesis in 2001.[38] But the facts do not support this hypothesis. As Rosovsky and Hartley explain:

> [G]rade inflation began in the 1960s when poor and minority students represented only 8 percent of the total student population. Furthermore, fully 60 percent of these students attended historically black colleges. . . . Most important, William Bowen and Derek Bok have demonstrated that, on average, black students in their sample did somewhat less well in college than white students who entered with the same SAT scores.[39]

More plausible than any explanation in terms of race but still problematic is the hypothesis championed by Biggs in "Fissures in the Foundation." She argues: "The fundamental cause [of grade inflation] was educators' loss of confidence in their ability to evaluate: the loss of confidence that resulted when their overinvestment in an ideal of scientific certainty betrayed them. This led to a loss of faith in their *right* to evaluate."[40] The principal difficulty with this hypothesis is that the "educators" who fit her description—disillusioned partisans of the psychometrics movement that flourished before 1950—were a small and specialized subset of the total set of college professors who were grading students in the late 1960s and early 1970s. It is telling that nearly all of the articles she cites from the formative period 1950–1968 were published in educational journals that most faculty members never read. Ph.D.s in arts and sciences, business, engineering, and so on were not schooled in educational psychology, or assessment strategies, or the history of education unless their particular subdisciplines demanded it. Many knew next to nothing about the rise and fall of the psychometrics movement and cared less.

Ironically, it may have been diminishing *interest* rather than diminishing *confidence* in the evaluation of students that helped tip grade inflation into an epidemic pattern. Increased birth rates, prosperity, and social mobility after the war produced a tidal wave of college applicants by 1960 that made it possible for many private and some public institutions to combine expansion in the size of their student bodies with greater selectivity. To teach this growing population of better qualified and more motivated students, colleges and universities hired an unprecedented number of young Ph.D.s and ABDs (all

but dissertation). Many of these neophyte professors were not much older than their students, and most recognized that their success in academe would depend on their productivity as scholars rather than their diligence as teachers. They soon discovered the fundamental law of grade/time management: lax grading is faster than stringent grading. They learned that students never challenge an A, and most will accept a B, but a grade of D or F needs to be explained in detail and may lead to time-consuming protests. Whatever the precise mix of causes, grading reached a tipping point by 1974 from which it could not right itself to the roughly normal curve and 2.3–2.5 GPA range that had prevailed for decades.

What happened to grades after 1978? To some extent, it depends on whom you ask. Adelman's studies of postsecondary transcripts report the following for national samples of high school students who earned ten or more credits in postsecondary education:

- For 1972 twelfth graders, transcripts for 12,600 students gathered in 1984 showed an average GPA of 2.70. Those who earned a BA or higher had an average GPA of 2.94.
- For 1982 twelfth graders, transcripts for 8,400 students gathered in 1993 showed an average GPA of 2.66. Those who earned a BA or higher had an average GPA of 2.88.
- For 1992 twelfth graders, transcripts for 8,900 students gathered in 2000 showed an average GPA of 2.74. Those who earned a BA or higher had an average GPA of 3.04.[41]

Arthur Levine and Jeanne Cureton's 1998 study of data from undergraduate surveys of 4,900 college students from all types of institutions in 1969, 1976, and 1993 found that from 1969 to 1993 As increased from 7 percent to 26 percent of all grades, while Cs decreased from 25 percent to 9 percent.[42] George Kuh and Shouping Hu's 1999 study comparing the GPAs of 52,000 students (half from the mid-1980s and half from the mid-1990s) *on a 5-point scale* found that the average had risen from 3.07 in the mid-1980s to 3.34 in the mid-1990s.[43] The U.S. Department of Education's *Profile of Undergraduates in U.S. Postsecondary Institutions: 1999–2000,* based on information obtained from more than 900 institutions on approximately 50,000 undergraduates, reported an average GPA of 2.9 for the nation as a whole during the period 1999–2000.[44] Stuart Rojstaczer's

Web site, www.gradeinflation.com, using data from the twenty-two colleges and universities that "have either published their data or sent their data to the author on GPA trends over the last 11 years," finds grade inflation remained relatively flat from the mid-1970s to the mid-1980s but has been climbing at a steady rate ever since. Rojstaczer reports an average GPA of 3.09 for the period 2001–2002.[45] The *National Survey of Student Engagement/Annual Report 2004*, based on a survey of 163,000 randomly selected freshmen and seniors at 472 four-year colleges and universities, found: "About two-fifths of all students reported that they earned mostly A grades, another 41% reported grades of either B or B+, and only 3% of students reported earning mostly Cs or lower."[46]

What lessons can we learn from this untidy set of national studies? What, if anything, do they tell us about the course and consequences of grade inflation in the twentieth century? Although no studies of comparable size or metrics are available for the period before 1972, small-scale surveys[47] and articles on grading practices such as those cited by Biggs, as well as anecdotal reports of normal curve grading, suggest that grading in the period 1930–1960 approximated the stability that Juola verified for Michigan State University from 1941 to 1965: fluctuations in the range from 2.3 to 2.5. In other words, it is reasonable to suppose that the average GPA in American colleges and universities hovered around 2.4 for at least three decades until grades began to rise in the 1960s. If we subtract this 2.4 estimate from the 2.9 reported in the Department of Education's *Profile of Undergraduates in U.S. Postsecondary Institutions: 1999–2000*, the implication is that the average GPA for the period 1999–2000 is now a half-point higher than it was on average from the period 1930–1960 and a tenth of a point higher than it was at the peak of the grade surge from the period 1968–1974. If this number is correct, then the average rate of increase between 1960 and 2000 on a 4-point scale has been .125 grade point per decade.[48] For price inflation, a comparable rate (3.125% per decade) would be delightfully benign, but since the harm done by grade inflation is cumulative, this rate is not good news for higher education.

Adelman's lower GPA figure of 2.7 implies an average rate of increase of .075 point per decade since 1960. A comparable rate for price inflation would be a measly 1.85 percent per decade—virtually no inflation at all. But again, this is a misleading comparison. To appreciate the cumulative harm of gradually rising grades, one needs

to compare them to a cumulative health problem such as obesity. Suppose that twenty-five-year-old males with a particular height and body type had an average weight of 170 pounds in 1960. Now suppose that the average weight for men in that category increased at an average rate of 1.85 percent per decade. This would mean that by 2000, the average weight of men in this category would be 182.58 pounds, a net gain of 12.58 pounds.

Various causes have been adduced to explain the gradual but unmistakably upward climb in grades since 1985. The political and ideological forces of the Vietnam War era are no longer relevant, and the political issues of the post-9/11 era do not seem to have affected grading one way or another. But competition for the recruitment and retention of students and internal policies that allow students to avoid being penalized by low grades are still significant factors. Other factors still contributing to grade inflation include increased reliance on adjuncts and part-time faculty as well as expanded use of student teaching evaluations for hiring, tenure, promotion, and salary decisions. Valen Johnson has been particularly persuasive in demonstrating that student grades bias student evaluation of teaching. In particular, he has discredited "the teacher effectiveness theory" that tries to explain the positive correlation between favorable grades and favorable evaluations as a double effect produced by good teaching.[49] Adjunct and junior faculty are especially vulnerable to student evaluation pressures, but even tenured full professors want to be liked and listened to. Since holding the line on grade inflation wins no applause from students and very little from colleagues, more and more faculty members have resigned themselves to the use of grades as multipurpose rewards rather than as indicators of performance.

WHY IS GRADE INFLATION HARMFUL?

To understand why grade inflation is harmful, one needs to see both that its harm is cumulative and that what it harms *most directly* is the capacity of grades to provide meaningful information to students. A slow, steady rise in prices is not a bad thing for economic health, even if it goes on forever, but a slow, steady rise in grades will make them dumb and dumber. As Fs, Ds, and even Cs become increasingly rare, the capacity of a grading system to convey precise information about student performance is diminished. Like "Newspeak" in Orwell's *1984*,

a grading system whose vocabulary is reduced to A and B cannot express the kind of critical distinctions students need to hear and teachers are uniquely positioned to make. Conflated grades can deny students information that could be useful for choosing majors and careers and deprive employers and graduate programs of a counterbalance to standardized test scores or reliance on the cruelly elitist criterion of institutional prestige.

Grade inflation also creates dissonance between the capacity of grades to inform and their power to reward. The irony of grade inflation is that it is driven by compartmentalized thinking about what grades really stand for. In a fascinating study, R. Eric Landrum found that "[s]tudents performing average work, who acknowledge themselves that their work is average, expect a grade of B or A more than 70% of the time, even though they realize that the grade for average work is C."[50] I have conducted similar studies in my classes and obtained compatible results. A-range grades, in particular, seem to retain their luster as badges of exceptional achievement, even when students know that half their classmates are getting the same badge.

Probably the worst results of grade inflation come from giving too many As and too few Fs. When A no longer distinguishes outstanding from good, teachers lack a formal means to inspire and reward exertion toward academic excellence. When F is only given to students who drop out or fail to turn in their work, rather than to all students who fail to meet course objectives, then colleges adopt the practice of social promotion that has stripped high school diplomas of credibility. Today, thousands of college graduates are teaching children, providing critical social services, and even gaining admittance to graduate programs without having mastered college-level skills or knowledge. There is a grim joke that medical students like to tell: "What do you call the student who graduates from medical school with the lowest grades in his class?" The answer is: "Doctor." One *hopes* that medical schools do not award passing grades to students who have not learned enough to practice their chosen profession, but there is little doubt that some undergraduate programs in fields as vital to public health as K–12 teaching do precisely this.

Another kind of harm arises from variations in grading leniency among teachers and disciplines. Other things being equal, teachers who grade more stringently than their colleagues are likely to attract fewer students and receive less favorable student evaluations. Johnson finds

that "within the same academic field of study, students are about twice as likely to select a course with an A– mean course grade as they are to select a course with a B mean course grade."[51] Moreover, disciplines such as mathematics and the natural sciences that have well-earned reputations for rigorous grading scare off students who could benefit from study in these areas but are fearful of putting their GPAs on the line. Johnson estimates that if differences in grading policies between divisions at Duke were eliminated, undergraduates would take about 50 percent more courses in natural science and mathematics.[52] Although the national consequences of grading inequities are extremely difficult to quantify, Johnson raises some disturbing questions:

> How does one measure the costs of scientific illiteracy? How has public discourse on issues ranging from stem cell research to genetic alteration of food products to discussion of missile defense technology to environmental protection policies been affected?[53]

Finally, there is a kind of harm that I have experienced firsthand but have not seen documented in the literature: obsessive and aggressive grade chasing by students and sometimes parents. The legendary "college gentleman" who considered any grade higher than C a stain on his escutcheon never sat in any of my classes. All the students I have known have preferred higher to lower grades—even if they were reluctant to work for them. What seems to have changed in recent years is the intensity of students' emotional investment in grades. I see lackluster students contesting grades as high as a B+. I hear students begging for extra assignments to pull their course average up to the A level. As dean and department chair, I had to deal with lengthy appeals for grade changes based on an astonishing range of personal hardships and legalistic objections to grading procedures. In the spring of 2000, an adjunct in my department with a record of good evaluations, received alarmingly bad evaluations from students in one of his sections. When I asked him about this poor rating, he explained that he had given Bs to two music majors early in the semester, and that they were so angered at his temerity that they had organized the class against him. (About 70 percent of all grades awarded by our Music Department are A or A–.) To prove his point, he showed me an e-mail from one of the music majors urging her classmates to punish him. (Apparently, she did not realize that the instructor was on her e-mail list.) Why

such anger? I suspect student anger and anguish over grades is fueled by the foolish conviction that small differences in GPA will significantly affect career opportunities and by the canny perception that grades have become negotiable multipurpose rewards rather than stable indicators of performance. The squeaky wheel gets the higher grade.

WHOM DOES GRADE INFLATION AFFLICT?

In one sense, grade inflation is pandemic; in another sense, epidemic. To the best of my knowledge, every American college and university that existed in 1960 has higher grades now than it did in 1960. On the other hand, there are noticeable differences among different types of institutions and larger differences among individual institutions. According to the *Profile of Undergraduates in U.S. Postsecondary Education Institutions, 1998–2000*, the national distribution for undergraduate GPAs at public doctorate-granting universities in the period 1999–2000 formed a kind of flattop bell curve with 9.9% of GPAs below 1.75, 9.6% above 3.75, and 52.5% between 2.25 and 3.24. In contrast, the national distribution for public two-year colleges in the period 1999–2000 was nearly flat from end to end, with 19.3% of GPAs below 1.75 and 16.6% above 3.75.[54] According to Rojstaczer's (somewhat ragged) data, Brown, Carleton, Columbia, Harvard, Harvey Mudd, Pomona, Princeton, Northwestern, and Williams had average GPAs of 3.30 or higher in 1998, while Hampden-Sydney, University of Houston, Norfolk State, Northern Michigan University, Sam Houston State, and State University of New York—Oswego had average GPAs in 1998 below 2.7.[55] This means that after thirty-eight years the latter institutions were still fairly close to the 2.4 midpoint that prevailed before 1960, while the former institutions were nearly a full grade point higher. A larger group of institutions tracked by Rojstaczer (about sixty-five) fell between these two extremes. This distribution is typical of epidemics (AIDS, obesity, and drug use hit some communities harder than others), but it also points to the high correlation between admissions selectivity and grade inflation.

The universities featured in popular exposés about grade inflation are almost always Harvard, Princeton, Dartmouth, and other bastions of privileged selectivity. Part of the reason for this narrow coverage is journalistic selectivity, or what Adelman calls "putting on the glitz."[56] The trials of the rich and famous attract more readers than the troubles

of their less-celebrated cousins. Yet behind the media hype is a pattern of real significance. With few exceptions, the colleges and universities that award the highest grades also are the institutions that reject the highest percentage of applicants and enroll students with the highest SAT or ACT scores and highest rankings in their high school classes. In other words, there is a strong correlation between highly selective admission practices and grade inflation. While highly selective admissions may not make schools rich and famous, it does make them elite. And this, in turn, suggests that warnings about the dangers of grade inflation may be exaggerated. Why should the practice of awarding better grades to better students be cause for national alarm? To give this question its due, we need to consider the viewpoint of scholars who question the existence of grade inflation as a nationwide phenomenon.

Clifford Adelman has sometimes denied the existence of grade inflation "at most institutions."[57] At other times, he has denied (perhaps inconsistently) that we can know whether grade inflation exists.[58] His denials and doubts rest chiefly on three arguments. As already noted, his postsecondary transcript studies show a decline in national average GPA from 2.70 for students who were twelfth graders in 1972 to 2.66 for those who were twelfth graders in 1982 and a rise to 2.74 for those who were twelfth graders in 2000. He summarizes the significance of these findings by saying: "The first—and major—point . . . is that, judging by both distribution of grades and GPAs, changes have been minor and complex since the high school student of 1972 went to college."[59] Complex perhaps, but not so minor. The chief difficulties with Adelman's contention that the rise in average GPA has been "minor" are: (1) 2.74 is significantly higher than the 2.4–2.5 that available evidence indicates prevailed for decades before the late 1960s; (2) the average GPAs for students in Adelman's most recent study who earned any kind of academic degree, even certificates, ranged from 2.96 to 3.04—a B average; (3) other national studies show even higher numbers; and (4) the harm done by grade inflation is cumulative.

Adelman's second argument is that if grades have gone up anywhere they have gone up at elite institutions. He has been quoted as saying: "They don't have articles about GPAs at Long Island University (LIU) or Montclair State."[60] This is misleading. Many non-glitzy institutions such as LIU and Montclair State *are* wrestling with grade inflation issues, even if their troubles are not reported in the *New York Times* or

the *Wall Street Journal*. My home institution, the College of New Jersey, formerly Trenton State College and a sister institution to Montclair State, had a mean GPA of 3.2 in the spring of 2004, and grade inflation is being debated on our campus—as it is at the University of Delaware, Ripon College, Loyola College in Maryland, Georgia Institute of Technology, and elsewhere. The core of truth in Adelman's second argument is that "elite" institutions, in the sense of having selective admissions, are far more likely to have high grading patterns than institutions that welcome all students. Still, it does not follow that grade inflation is not epidemic because it affects some institutions more than others. Public health authorities warn that obesity has reached epidemic proportions in the United States since 30 percent of Americans are obese. It would be foolish to reject their warning on the grounds that 70 percent of Americans are not obese.

Adelman's third argument is conceptual rather than statistical. He has been quoted as saying "In the matter of grades, it's impossible to judge inflation. . . . For every class in which it is claimed that grades have risen, I will need [proof] that there has been inflation . . . I want the same assignments, with same criteria, with same people judging them over 30 years. Do you think that's possible?"[61] I suppose the intent of this argument is to challenge any attempt to prove that significant grade inflation exists, but its salutary (if unintended) effect is to expose the flaw in defining grade inflation as an upward shift in the grade point average of students over an extended period of time *without a corresponding increase in student achievement.* I agree with Adelman, that attempts to make the case for grade inflation in this sense run into insurmountable obstacles, though, unlike him, I think the solution is to advance a better definition of grade inflation rather to question its reality.

The evidence most often cited for a growing gap between grades and achievement comes from standardized college entrance exams and the proportion of students requiring remedial education. As Rosovsky and Hartley explain, the average combined SAT scores declined 5 percent between 1969 and 1997, and the proportion of students requiring remedial education increased between 1987 and 1997.[62] Yet what these numbers indicate is a decline in preparedness for college education. They do not preclude the possibility that, despite weaker preparation, students during these years achieved more when they got to college. Neither do these numbers tell us much about institutions that became more academically selective during the years in question—

the institutions where GPAs have risen most dramatically. Again, preparedness and achievement are not the same thing. Recruiting students with top SAT or ACT scores and high school ranks does not justifying awarding them top grades: students should not be graded on their accomplishments before coming to college but on the work they do after being admitted.

To make the case *either for or against* a growing gap between grades and achievement, one would need to compare grades with measures such as: (1) students' "objective mastery" of a discipline; (2) the measurable superiority of current students' work over the work of students in former years; or (3) the measurable superiority of student work at one institution to the work of students at other institutions. Yet a little reflection on the practicality of making such measurements displays the implicit wisdom in Adelman's third argument.

To begin with, most disciplines are unable to agree upon measurable standards of objective mastery beyond basic comprehension, rote knowledge, and threshold skills. The notion of "grades as an objective measure of mastery" has limited application in postsecondary education. The achievement of native proficiency is a clear and testable measure of mastery in learning to speak a foreign language, but there are few parallels elsewhere in higher education. One does not master Milton, Mozart, or metaphysics.

Comparing the work of current students to that of former students has superficial appeal but is woefully impractical. Recently hired colleagues cannot make such comparisons, and longtime colleagues are unlikely to have sufficiently reliable memories. Perhaps one could compensate for these obstacles by persuading the faculty to spend some time every semester rereading old tests and papers, but I do not think faculty could or should be persuaded to use their time in this way. Besides, as fields of study change, so do our expectations for students. The level of skills and knowledge expected in a genetics course today is and ought to be considerably higher than it was twenty years ago.

Comparing the work of one's own students with that of students at other institutions is similarly impractical, except for those rare departments that require and trust standardized tests.[63] Few professors make any effort to find out how students at other institutions are performing, but even if more did and dependable comparisons were established, their discoveries would not provide decisive guidance in grading. If Harvard professors, for example, graded their undergraduates

in strict comparison to national averages of student achievement, then Harvard would probably give nothing but As. Of course, Harvard could pick a tougher standard of comparison, say other Ivy League schools, or all Research I universities, or the schools on Barrons's most selective list. But nothing inherent in the functions of grading would favor one choice or another, unless Harvard picked schools (if such could be found) where student achievement was precisely on par with Harvard's. The labor involved in verifying this comparative "same as Harvard" standard would be considerable, but more importantly it would be wasteful, since the standard that matters is the achievement of students who are currently *at* Harvard. If an A on a Harvard transcript signified that a student's achievement in a course was outstanding (and not merely above average) *in comparison with her classmates*, then it would say something of real significance.

In summary, then, Adelman's third argument against grade inflation is persuasive if we construe it as a case against definitions of grade inflation that rely on hard-to-make comparisons of student achievement across time and space rather than as a case against grade inflation itself.

Alfie Kohn also has taken up the cudgel against educators who worry about high grades. He argues with considerable flair that higher grades are not inflated unless they are undeserved, and that has never been demonstrated.[64] He suggests that there may be alternative explanations for higher grades at selective institutions: "Maybe students are turning in better assignments. Maybe instructors used to be too stingy with their marks. . . . Maybe the concept of assessment has evolved, so that today it is more a means for allowing students to demonstrate what they know."[65] "The real question," he suggests, "is why we spent so many years trying to make good students look bad."[66]

The key notion here is "deserving a grade," which is paired in turn with "turning in better assignments," "less stingy professors," and an "evolved concept of assessment." It sounds reasonable to say "if every student in a class deserves an A, then every student in the class ought to receive an A." But whether a student deserves an A depends *both* on what that student does (as ascertained by assessment) *and* what an A stands for. If an A stands for "near the top of the class," then it will not be possible for every student in the class to *deserve* an A. The question of what grades should stand for is complex and must be answered with an eye to a variety factors, such as estab-

lished associations, the portability of grades from one institution to another, grade inflation, and so on, but it cannot be settled by assessment. The task of assessment is to measure or otherwise ascertain a student's performance. The task of grading is to assign a grade to that performance that conveys true and useful information to the student and to others who are granted access to the grade.

When *Consumer Reports* rates self-propelled lawn mowers, it assesses each model on a variety of performance measures—evenness, mulch, bag, side discharge, handling, and ease of use—and then ranks each model in terms of overall score with ratings of "excellent," "very good," "good," "fair," or "poor" for every performance category. There is no preordained relationship between these ratings (think of them as "grades") and the functions measured, functions such as the distribution of x clippings per square foot. The technicians at Consumer Union have to make grading decisions that provide the readers of *Consumer Reports* with useful information. They cannot discover what rating to assign "mulch" by measuring the clippings a little more carefully or creatively. Moreover, it is clear what would happen to the readership of *Consumer Reports* if the editors, fearful of making good lawn mowers look bad, replaced rankings and ratings with summary grades that deemed half the models "excellent" and another 40 percent "very good." Even if *all* of the models tested were fine choices, readers would still want information on relative performance in various categories.

Allowing a student to demonstrate what she knows may be a good strategy for teaching and assessment, but it will never entail that she *ought* to receive a particular grade for her performance. Assessment does require the use of standards, but these are standards against which performance is measured: vocabulary memorized, equations solved, interpretations defended, and so on. Grades do not follow from the application of such standards, unless one has *already built them in.* Kohn's appeal to an "evolved concept of assessment" as a justification for giving high grades simply begs the question.

The deepest confusion in Kohn's diatribe against "the dangerous myth of grade inflation" is his insistence that using balanced grade distributions to motivate and improve student learning is counterproductive. He thinks grades themselves are "a threat to excellence"[67] and asks: "How can professors minimize the salience of grades in their classrooms, so long as grades must still be given?"[68] What he fails to see is that inflated grades make students more, not less, grade conscious

and at the same time reduces their incentive to work harder. One of the virtues of normal curve grading was that grades of B and C did not demean students; these grades indicated that a student had met the requirements of the course and had done so as or more successfully than most of the students in the class—though not as well as the students who received the grade reserved for top performers. Today a B or C often indicates that a student has disappointed the teacher and, therefore, has been excluded from sharing in the cornucopia of As available to every good student in the class. Thus Bs and Cs have become marks of disapproval and A's counterfeit tokens of excellence. At the same time, As no longer motivate diligent students to excel, and lazy students find they can pass a course (often in the B range) merely by coming to class and turning in assignments. We tell students they should spend two hours on academic work outside of class for every hour they spend in class (roughly thiry hours of outside work for a full-time student), but the *National Survey of Student Engagement/ Annual Report 2005* found that only 8 percent of first-year students and 11 percent of seniors said they spent more than twenty-six hours a week preparing for class.[69]

I am half-tempted to agree with Kohn that higher education might be better off if we could get rid of grades altogether and replace them with written assessment and evaluation reports. My hesitation is twofold. First, unless written reports are standardized in some way, they are likely to be ignored by graduate programs and prospective employers. As it is, too little weight is given to the judgment of teachers and too much to objective examinations or institutional reputation. Second, letters of recommendation are the only example we have of written assessments as a nationwide practice in higher education, and the record here is dismal. Rosovsky and Hartley observe: "What evidence is available—empirical, anecdotal, and experiential—leads us to conclude that letters of recommendation suffer from many of the same, or worse, weaknesses and problems as grades."[70]

Rather than abandon grades altogether, I think we should put them in their place. Grades are a blunt instrument. They are a handful of symbols that can be effective in communicating limited information to students (and with student permission to others) if we do not burden them with ambiguous or incompatible messages. Normal curve grading worked well, not because it was hard on students, but because it was clear to students. Perhaps we can do better than that; we have certainly

done worse. The practical question is how to get the current epidemic of grade inflation under control and steer American higher education toward constructive grading practices. I deal with that question elsewhere in this volume.

NOTES

1. Henry Rosovsky and Matthew Hartley, "Evaluation and the Academy: Are We Doing the Right Thing?: Grade Inflation and Letters of Recommendation," Occasional Paper (Cambridge, MA: American Academy of Arts and Sciences, 2002), 21. This well-crafted study of inflated grades and letters of recommendation reflects the mainstream view of educators who believe that grade inflation is a serious problem and needs to be addressed.

2. Typical of this view is David Basinger, "Fighting Grade Inflation: A Misguided Effort," *College Teaching* 45 (Summer 1997): 81–91.

3. Clifford Adelman, *The New College Course Map and Transcript Files: Changes in Course-Taking and Achievement, 1972–1993* (Washington, DC: U.S. Department of Education, 1995), viii, 266.

4. Clifford Adelman, "As Aren't That Easy," *New York Times* (May 17, 1995), p. A19.

5. Clifford Adelman, "Putting on the Glitz: How Tales from a Few Elite Institutions Form American's Impressions about Higher Education," *Connection: New England's Journal of Higher Education* 15:3 (Winter 2001): 25.

6. Clifford Adelman, *Principal Indicators of Student Academic Histories in Postsecondary Education, 1972–2000* (Washington, DC: U.S. Department of Education, Institute of Education Sciences, 2004), 77.

7. Alfie Kohn, "The Dangerous Myth of Grade Inflation," *The Chronicle of Higher Education* (November 8, 2002): B7.

8. In an influential article published in 1985, Louis Goldman stated: "Grade inflation is the term used to describe an upward shift in the grade point average." See Goldman's "Betrayal of the Gatekeepers," *The Journal of General Education* 37:2 (1985): 98.

9. Rosovsky and Hartley, "Evaluation and the Academy," 3. Rosovsky and Hartley cite Goldman's article without comment, but they add the qualification "without a corresponding increase in student achievement." Kohn defines grade inflation as "an upward shift in student's grade point averages without a similar rise in achievement." See Kohn, "The Dangerous Myth of Grade Inflation," B7.

10. Mary Biggs and I originally suggested "grade conflation" as an alternative name for grade inflation, but I now believe grade inflation should be regarded as a species of grade conflation. See Richard Kamber and Mary Biggs, "Grade Conflation: A Question of Credibility," *Chronicle of Higher Education* (April 12, 2002).

11. Iver Peterson, "Flunking Is Harder as College Grades Rise Rapidly," *New York Times* (March 13, 1972), pp. 1, 2. (My thanks to Iver Peterson for providing me with a copy of this article.)

12. William A. Sievert, "Lax Grading Charged at California Colleges," *The Chronicle of Higher Education* (November 27, 1972): 3.

13. "Too Many A's," *Time* (November 11, 1974): 106. This article makes no mention of "grade inflation." It says: "College grades along with almost everything else have been going up lately. . . . Indeed, in the past few years the grade glut has been spreading across academe."

14. Robert J. Staff and Richard B. McKenzie, *An Economic Theory of Learning: Student Sovereignty and Academic Freedom* (Washington, DC: U.S. Department of Education, 1973). This is the final report of a grant sponsored by the National Institute of Education (DHEW). It presents analogies between both price and monetary inflation. It also compares faculty to the Federal Reserve.

15. David Riesman, *The Lonely Crowd: A Study of the Changing American Character* (New Haven, CT: Yale University Press, 1950), 91.

16. For a more detailed version of this analysis, see Richard Kamber and Mary Biggs, "Grade Inflation: Metaphor and Reality," *The Journal of Education* 184:1 (2004): 31–37.

17. Some institutions have experimented with super grades that are higher than the regular top grade in their grading systems. Johns Hopkins University, in the early 1960s, had an H, which stood for Honors, and had a value of 5 points as opposed to 4 points for an A. Stanford University has for twenty-five years awarded 4.33 points for an A+. Arizona State University began giving the weighted A+ in the fall of 2004. See Ramin Setoodeh, "Some Collegians Make Grade with A-Plus," *Los Angeles Times* (February 4, 2004): 96. Super grades of this kind raise the upper limit in a grading system, but they do not remove that limit. In theory, a grading system could mimic the capacity of prices to rise without limit by allowing faculty members to invent higher grades at will. At Euphoric State University, professors could exceed the institution's regular top grade of 4 by giving a 5, 6, 7, or any higher number they believed was justified. To the best of my knowledge, this has never been done.

18. Valen Johnson remarks: "In a very real sense, professors pay grades to students for mastery of course material, and students barter these grades for jobs or entrance into professional or graduate school." See Valen E. Johnson, *Grade Inflation: A Crisis in College Education* (New York: Springer-Verlag, 2003), 196. If the "real sense" he has in mind is that grades do *in fact* serve as incentives and rewards for learning, then his remark is unobjectionable. But if he means that grades *should be* regarded as means of exchange rather as tokens of recognition, then this remark seems inconsistent with his well-reasoned efforts to combat grade inflation. For example, he recommends reforms aimed at reducing grade inflation in order to rectify the imbalance between enrollments in natural science

and mathematics courses and enrollments in the humanities and social sciences. But if grades are pay, then it would be more reasonable and practical to urge natural science and mathematics departments to stop being so cheap and pay their students a competitive rate.

19. Paul A. Samuelson and William D. Nordhaus, *Economics*, fourteenth ed. (New York: McGraw-Hill, 1992), 596.

20. Mary Lovett Smallwood, "An Historical Study of Examinations and Grading Systems," *Early American Universities: A Critical Study of the Original Records of Harvard, William and Mary, Yale, Mount Holyoke, and Michigan from their Founding to 1900* (Cambridge, MA: Harvard University Press, 1935), 8.

21. Frederick Rudolph, *The American College and University: A History* (New York: Random House, 1962), 47.

22. Ibid.

23. Rudolph, *The American College and University*, 48.

24. Mark Hopkins, quoted in Frederick Rudolph, *Mark Hopkins and the Log: Williams College, 1836–1872* (New Haven, CT: Yale University Press, 1956), 221.

25. Rudolph, *The American College and University*, 282.

26. Report of the Committee on Raising the Standard, 1894, Harvard University archives. (My thanks to Matthew Hartley for providing me with this citation.)

27. Arvo Juola, "Grade Inflation in Higher Education: What Can or Should We Do?" *Education Digest* 129 (1976): 917.

28. Ibid., 2.

29. Ibid.

30. Ibid., 3.

31. Ibid., 2.

32. Arvo Juola, "Grade Inflation—1979. Is It Over?," *Education Digest* 189, (1980): 129. The 1979 survey Juola used to complete this study did not query the same population of institutions as his first two studies. He used a systematic sample of 361 institutions offering graduate degrees, and received usable responses from 180. Rosovsky and Hartley state incorrectly that Juola's surveys "found that grade inflation continued unabated between 1960 and 1977." See Rosovsky and Hartley, "Evaluation and the Academy," 5.

33. Ibid., 11.

34. See, for example, Rosovsky and Hartley, "Evaluation and the Academy," 7–8; Arthur Levine, *When Dreams and Heroes Died: A Portrait of Today's College Student, A Report for the Carnegie Council on Policy Studies in Higher Education* (San Francisco: Jossey-Bass, 1980); James B. Twitchell, "Stop Me Before I Give Your Kids Another A," *Washington Post* (June 4, 1997).

35. For a detailed account of these practices, see Herbert J. Riley et al., *Current Trends in Grades and Grading Practices in Undergraduate Higher Education: The Results of the 1992 AACRAO Survey* (Washington, DC: American Association of Collegiate Registrars and Admissions Officers, 1994).

36. David Riesman, *On Higher Education* (San Francisco: Jossey-Bass, 1980), 78.

37. Ibid., 80.

38. Harvey Mansfield, "Grade Inflation: It's Time to Face the Facts," *The Chronicle of Higher Education* (April 6, 2001): B24.

39. Rosovsky and Hartley, "Evaluation and the Academy," 8. William G. Bowen and Derek Bok, *The Shape of the River* (Princeton, NJ: Princeton University Press, 1998).

40. Mary Biggs, "Fissures in the Foundation," Chapter 7 in this book.

41. Grading data for the three transcript studies are summarized in Adelman, *Principal Indicators of Student Academic Histories in Postsecondary Education, 1972–2000*. The two earlier reports are: (1) Adelman, *The New College Course Map and Transcript Files: Changes in Course-Taking and Achievement, 1972–1993* viii, 266, and (2) Clifford Adelman, Bruce Daniel, and Ilona Berkovits, *Postsecondary Attainment, Attendance, Curriculum, and Performance: Selected Results from the NELS: 88/200 Postsecondary Education Transcript Study* (Washington, DC: U.S. Department of Education, National Center for Education Statistics), 27.

42. Arthur Levine and Jeanne Cureton, *When Hope and Fear Collide: A Portrait of Today's College Student* (San Francisco: Jossey-Bass, 1998).

43. See George D. Kuh and Shouping Hu, "Unraveling the Complexity of the Increase in College Grades from the Mid-1980s to the Mid-1990s," *Educational Evaluation and Policy Analysis* 21:3 (1999): 297–320.

44. Laura Horn, Katharin Peter, and Kathryn Rooney, *Profile of Undergraduates in U.S. Postsecondary Education Institutions, 1999–2000* (Washington, DC: NCES, 2002), 168. The printed version of this document reports grades descriptively (e.g., "mostly As"), but quantitative GPAs are available through the Data Analysis System Web site, http://www.nces.ed.gov/das/.

45. Stuart Rojstaczer, "Grade Inflation at American Colleges & Universities" http://www.gradeinflation.com/, accessed March 17, 2003, 1, 3.

46. Center for Postsecondary Research, *National Survey of Student Engagement/ Annual Report 2004* (Bloomington: Indiana University, School of Education, 2004), 13.

47. A good example of a small-scale survey prior to the 1970s is M. J. Nelson, "Grading Systems in Eighty-nine Colleges and Universities," *Nation's Schools* 5 (June 1930): 67–70. See Mary Biggs's analysis of this report in "Fissures in the Foundation," Chapter 7 in this book.

48. Rojstaczer estimates the average rate of increase over the past thirty-five years at .15 points (out of 4.0) per decade, 3.

49. Johnson, *Grade Inflation*, 1–131.

50. R. Eric Landrum, "Student Expectations of Grade Inflation," *Journal of Research and Development in Education* 32:2 (Winter 1999): 124–28, 126.

51. Johnson, *Grade Inflation*, 192.

52. Ibid.

53. Ibid., 194.

54. *Profile of Undergraduates in U.S. Postsecondary Education Institutions, 1998–2000*, Table 2.3.

55. Rojstaczer, "The Data Currently Available," http://www.gradeinflation.com/, accessed

56. Adelman, "Putting on the Glitz," 24.

57. Adelman, "A's Aren't That Easy," p. A19 and *The New College Course Map and Transcript Files*, viii, 266.

58. See comments by Adelman quoted in "The Great Grade-Inflation Lie," *Boston Phoenix* (April 23–30, 1998), and "What's in an 'A' Grade?," *Ithacan Online* (March 2, 2000).

59. Adelman, *Principal Indicators of Student Academic Histories in Postsecondary Education, 1972-2000*, 77.

60. Adelman, "The Great Grade-Inflation Lie," 2.

61. Adelman, "What's in an 'A' Grade?," 2.

62. Rosovsky and Hartley, "Evaluation and the Academy," 6.

63. The Graduate Record Examination Subject Tests, the National Accountancy Examination, and the Test of English as Foreign Language are examples of examinations that might serve this purpose.

64. Kohn, "The Dangerous Myth of Grade Inflation," B8.

65. Ibid. Kohn may be reacting against what Johnson calls "the traditional view of grading," namely, the view that "grades are devices used by faculty to maintain academic standards and to provide summaries of student progress" (see Johnson, *Grade Inflation*, 3), since Kohn is skeptical about the usefulness of grades to maintain standards. If so, he is giving this view more weight than it deserves. Traditional or not, this view mixes the fundamental purpose of grades (to provide summaries of student progress—or as I prefer to state it—to provide students with information about their performance) with a derivative use (to maintain academic standards). Grades may or may not serve the latter, but they must serve the former.

66. Ibid.

67. Kohn, "The Dangerous Myth of Grade Inflation," B9.

68. Ibid.

69. Center for Postsecondary Research, *National Survey of Student Engagement/ Annual Report 2005* (Bloomington: Indiana University, School of Education, 2005), 43.

70. Rosovsky and Hartley, "Evaluation and the Academy," 16.

This page intentionally left blank.

CHAPTER FOUR

Grade Inflation and Grade Variation: What's All the Fuss About?

HARRY BRIGHOUSE

In 2001, Harvard history professor Harvey Mansfield published an article in the *Chronicle of Higher Education* criticizing his colleagues at elite universities for giving their students high grades.[1] To be fair, he targets himself with the same criticism; he believes that he too participates in this bad practice, which he dubs, in the title of his piece, "Grade Inflation." This might expose him to charges of hypocrisy. But he is not necessarily wrong to give out higher grades than he thinks are generally proper. In fact, my own practices in my first years of teaching at university followed roughly Mansfield's trajectory. I came to American higher education from an environment in which only truly excellent work was rewarded with high grades or praise. To give the reader an idea, only 5 percent of my class in an English university were awarded first class honors degrees. Throughout my high school career, I was regarded as a strong student, and I consistently attained Bs. There was no shame in that; there were very few As.

But I found a very different environment in American higher education. Students choose courses; very few are compulsory, even within a major. The courses are, frankly, hard. I assign difficult material and expect students to read it and to put a great deal of work into their written assignments. My exams are not exactly unfair—I give students fair warning of what is coming up and what kinds of questions will be asked, and advice on how to prepare. But they are not easy—students have to master a lot of material and be prepared to display that mastery.

What happened? Initially I assigned grades without regard to how other people did it. Since I have high standards (in some artificial sense), that resulted in very few As; very few indeed. So when I

learned how routinely As were given out in other courses in the humanities, I felt that I was actually treating my students unfairly. They were being penalized in terms of something that mattered to them, and other people, for taking the demanding course. This seemed just wrong. So I became more generous, and with As in particular.

In an environment in which you believe that students are generally awarded higher grades than they should be, it can be morally appropriate to do the same oneself, because in the absence of some collective action to depress grades, it is unfair to penalize students for taking your course.

In the final section of this chapter I shall propose a reform that I believe would inhibit grade inflation (if, indeed, there is any such thing as grade inflation) but that is also desirable on numerous other grounds. But before making my proposal I want to examine what constitutes grade inflation, ask whether there is any such thing, and deflate some of Mansfield's criticisms of his colleagues and his university. I shall also look at the evidence concerning grade variation, which I think is a more troubling phenomenon.

GRADE INFLATION AND THE PURPOSES OF GRADING

What is grade inflation? A simple definition, and the one that I shall deploy here, is that grade inflation occurs when, over time, mean grades increase faster than the mean quality of work produced by students, or mean grades decrease more slowly than the mean quality of work produced by students.

Let us assume for the moment that grade inflation is a real phenomenon. Why would it be objectionable? We can only answer this question by reference to the appropriate purposes of grading, so let's start by looking at those.

I can discern three widely accepted central purposes of grading.

1. Grades inform students about the quality of their own performance. Students want to know, and have a legitimate interest in knowing, whether their performance conforms to standards it is reasonable to expect from them at the current stage in their intellectual development. This is especially important in a system like the U.S. higher education system, in which stu-

dents are simultaneously pursuing studies in several disparate classes and disciplines, and in which they therefore need information on which to base their decisions about budgeting their time and effort from week to week.[2]

2. Grades inform future employers, vocational schools, and graduate schools about the quality of the student. One thing employers are interested in is how well the applicant has applied herself to demanding tasks; another is how capable she is of performing those tasks well. Grades aggregate this information for the employer and thus help him sort applicants. The information grades give employers is crude; it is impossible to disagreggate the effort from the talent, and it is difficult to compare grades across disciplines and, even more so, between institutions.
3. Finally, grades are pedagogical tools for eliciting better performance from the student. This purpose, unlike the others, is highly individualized. So we might encounter a student, Celia, who lacks self-confidence and is easily discouraged. Receiving higher grades might encourage her to put more effort in and thus raise her performance (and thereby achieve what really matters, which is learning more). This motivation for higher grades is identified by Mansfield as the "self-esteem" motive.[3] Another student, Betty, might, in contrast, have an unjustified surfeit of self-esteem; she might be coasting, and a slightly lower grade would perform the equivalent function of eliciting more effort and, consequently, more learning.

These functions have a complicated relationship. The third is both parasitic on the first two and at odds with them. It depends for its success on the student believing that it is not being used, but that the other functions (which demand a nonindividualized grading strategy) are at work. If the third strategy is used too frequently and too openly it loses its effectiveness, because students lose confidence in the grade and therefore cease to respond to it in the way that the strategy requires. A second comment to make about the third function is that the conditions that prevail in higher education, especially in the larger universities, strongly militate against its effective use. For it to work the teacher has to know the student, have some sense of her intellectual abilities and the trajectory of her intellectual development, and be able to make reliable conjectures about her response to the grade.

I doubt that grade inflation makes a big difference to the first function. As long as students know that grades are inflated they know to discount the information they might get, and seek other information (they look at other students' work, or ask the professor for information, etc.). What matters for this function is not the individual professor's grading policies but the culture of grading into which the students have been inducted. An individual professor can have a grading policy that bucks the trend, and can explain that he does so, but this will not be helpful to the students who have internalized a different standard. This is one of the reasons it is important for an institution to pay attention to its grading practices and to have some degree of uniformity. If the mean grade goes up from a C to a B over time, then the informational value of the grade for the student is still strong.

Furthermore, for this function as for the others, it is not only the grade that performs it. We can inform students about how well they are performing by telling them, by writing marginal comments on their papers, by taking their comments seriously in class, and so on.

Why should the second function be affected by grade inflation? It is worth noting that even if, say, a B– were the lowest grade that any student received it would still be possible to discriminate between students on the basis of their entire GPA, which usually goes to two decimal places. Compression at the top would have to be much more severe than any of its critics actually claim before it became impossible to distinguish between students on this basis. Two students might be only 0.2 apart, rather than 0.35 apart, but if they are still ranked ordinally employers can see that. But in fact employers, graduate schools, and professional schools look at the candidate much more broadly than the story told by their grades. As long as we all know that grade inflation has happened (and we do) we discount the information provided by grades. But we would have to do this anyway, because our applicants come from diverse institutions which, we know, have very different qualities of intake and very different grading standards. Is a B from Stanford the same as a B from UW Stevens Point? I do not know. What about a C? A C is probably more similar than an A. But employers always look at other signals—letters of recommendation, breadth (or, if desired, narrowness) of course distribution, self-presentation, interviewing behavior, and the like.

Grade inflation would have its worst effect on the third function. Why? An overall inflation of grades affects the ability of the professor

to use the grade for motivational purposes, not least by enculturating students to receiving higher grades so that lower grades are more likely (than in a noninflated regime) to depress their self-esteem. If lower grades always depress self-esteem, then they cannot be used as easily to elicit effort, and if higher grades are the norm, then they cannot be used to encourage students.

So if there were grade inflation, that would be a bad thing, but only a moderately bad thing. As grades increase across the board, the ability of professors to use grading as a pedagogical and motivational tool is limited. But, of course, professors are a devious bunch and can, if so minded, deploy other strategies to manipulate their students into working harder and better; in losing this device, they lose only *one* device. Well-chosen comments, whether spoken or written, can be much more effective motivators for most students, and they would be even more powerful in a world in which grades were accorded less significance than they are in most contemporary universities.[4]

Mansfield does make a comment that might make this problem seem worse than it really is. He says that "Grade inflation compresses all grades at the top, making it difficult to discriminate the best from the very good, the very good from the good, the good from the mediocre. Surely a teacher wants to mark the few best students with a grade that distinguishes them from all the rest in the top quarter, but at Harvard that's not possible." I'm not sure why it is so important to hive off a handful of students for special recognition, if the whole top quarter is doing work that is very good. In nearly twenty years of teaching in research universities I have come across many students who are smarter than I am and more promising than I was at their age, but only four or five students whose work placed them unambiguously well above the rest of the top quarter, and only one whose work stunned me. Reserving an A (or A+ or A++) for them takes grades too seriously. How could the one stunning student know that he was being rewarded with a stunning grade? And why should he care?[5] A professor can reward, or "mark," those students' work much more effectively with verbal or written praise, or with a request to meet to discuss the paper, or with frank admiration of a thought in the public forum of the classroom. Only students unhealthily obsessed with their grades would be more motivated by a special grade than by alternative forms of recognition.[6]

Before looking at the evidence concerning grade inflation, I want to deflect another possible objection to it, which is analogous to one

of the best reasons for objecting to monetary inflation. One of the reasons monetary inflation is bad is that it effects a redistribution of resources from savers to borrowers. Borrowers capture a rent, and this is both inefficient (for the economy) and unfair (to the savers). Does grade inflation cause some similar redistribution? If so, it would have to be redistribution of opportunities, or of status, or of honor, from the earlier low-grade regime generation to the later, grade-inflated generation. But it is hard to see how any redistribution of job opportunities could occur. Grades are typically relevant to opportunities *within* a cohort; employers look at the grades of prospective entry-level employees because there is limited information about their prospective job performance, but once the employee has a work record, the employer or prospective employer discards information about grades in favor of that. Many readers of this book have been involved in academic hiring: What would you think if one of your colleagues asked to look at the undergraduate (or high school) grades of a prospective assistant professor? Or compared the graduate school grades of competing candidates for a senior position? Redistribution of status or honor from the older to the younger could, in theory, be prompted by grade inflation, but I think it very unlikely, for two reasons. First, as long as people believe there is grade inflation, they discount the successes of later cohorts in their allocation of honor and status. Second, post-graduation, honor, and status are garnered much more through achievement and income than through displays of one's pre-career in college. I really doubt that anyone is misled in the ways that would be necessary to cause an unfair redistribution of status or respect.

IS THERE ANY GRADE INFLATION?

So grade inflation would be a minor problem in the life of the university, and especially in the pedagogical organization of the professor, if it were a real phenomenon. But is it real? Mansfield offers no evidence at all that there is grade inflation. He does offer anecdotes about how it arose, but nothing more. When he first raised the issue of grade inflation, he said "that when grade inflation got started, in the late 60s and early 70s, white professors, imbibing the spirit of affirmative action, stopped giving low or average grades to black students and, to justify or conceal it, stopped giving those grades to white students as well."[7]

He excuses his lack of evidence for grade inflation and its initial triggers by pleading lack of access to the facts and figures. But his comments about what he lacks access to lead me to suspect that he does not understand what would constitute hard evidence of grade inflation. This is what he says:

> Because I have no access to the figures, I have to rely on what I saw and heard at the time. Although it is not so now, it was then utterly commonplace for white professors to overgrade black students. . . . Of course, it is better to have facts and figures when one speaks, but I am not going to be silenced by people [referring to a dean and the then-president, of Harvard] who have them, but refuse to make them available.[8]

The problem here is that "figures" tell us nothing about whether there is grade inflation, and that nobody in the administration has the "facts" that would tell us (or, if they do, Harvard is a unique educational institution). If grades have increased at Harvard, then Harvard is not alone. Stuart Rojstaczer has compiled data from selections of private and public universities, showing mean GPAs increased by .15 in both groups between 1991 and 2001.[9] Using data from the same universities, he finds an increase in mean GPA of .68 in private universities and .5 in public universities from 1967 to 2002. But this is simply no evidence at all of grade inflation. Why?

Let us suppose that the mean grade has increased dramatically during the time in which Mansfield has taught at Harvard. Does this mean that students are now being given higher grades for the same quality of work, or the same grades for lower-quality work, than before? No. The improved grades may reflect improved work. Perhaps the teaching at Harvard has improved since Mansfield was hired. Perhaps the students are more talented or harder working or better prepared on entry or all three. If the first of those possibilities sounds unlikely, think again: it is hard to imagine even a legacy student as weak as now-President George W. Bush gaining admission to Harvard or Yale or any other elite college today, as he did (to Yale, admittedly) in 1964. In the period we are discussing, the mean number of children born into elite families has declined, enabling those families to invest more in each child; Harvard also has expanded dramatically the talent

pool from which it draws, by admitting women (although the first woman was admitted in 1950, women were admitted on an equal basis with men only in 1973) and hiring professors in a much more open labor market than when Mansfield was hired; and more of those women have been socialized to be ambitious in academic and career terms over that time. The "Gentleman's C," which both the 2004 presidential candidates were awarded by Yale in the 1960s, is reputedly a thing of the past, and so are the gentlemen at Yale who were awarded them. How could we know whether there was grade inflation? To know it would require a large, year-by-year database of the actual work done by Harvard students (including, presumably, evidence of their classroom participation) and the matching grades. I would be very surprised if the administrators have such a database access to which they are guarding. Mansfield could know: if one existed, he would have been asked to contribute to it. But he does not even seem aware that that is what would be needed.

To illustrate the point that higher grades are not necessarily inflated grades, consider my own story. I took a two-year leave from the University of Wisconsin, Madison, and returned to find a marked improvement in the quality of work my undergraduate students were turning in (I know, because, unusually, I keep a not-very-systematic collection of past work). The grades I assigned reflected this improvement. I sought an explanation (the default being that my teaching had improved beyond recognition) and found (to my disappointment) that Madison had tightened its admissions requirements in the years prior to my departure; I was now seeing the fruit of those changes.

The assumption that higher grades are entirely due to grade inflation has a corollary that is, to put it mildly, rather surprising. The corollary is this: There has been no productivity gain in the education industry during the period under consideration.[10] This would be surprising, for at least three reasons. First, during the period under consideration, we know that other industries have experienced considerable productivity gains. Second, those gains have produced a job structure in which the level of education needed to perform the median occupation, and the average occupation, has increased considerably. So the economy seems to require a more educated workforce and seems to function productively; this should be impossible if there had been no productivity gain in education.[11] Third, setting aside the possibility that pre-college education has improved over this time, within

research universities at least there has been a decline in the number of courses each faculty member teaches. Research might have benefited from this, but it also is possible that teaching has benefited, especially if, as we who teach in these universities like to believe, there is some sort of synergy between teaching and research.

Mansfield attributes putative grade inflation to the response of professors to affirmative action and the self-esteem movement. In fact, there *is* a mechanism that you would expect to have triggered grade inflation in the period in question: the introduction of, and gradual accommodation to, student evaluations of teaching. Insofar as these evaluations are used for purposes of deciding pay raises, promotions, and tenure, and insofar as higher grades influence student evaluations, you might expect them to prompt some sort of grade inflation. Professors have a self-interested reason for trying to trigger better evaluations, and grades are not only an effective method of doing this, but they may be a more effective method than improving their teaching, since better teaching does not necessarily result in better evaluations.

Student evaluations of teaching are, indeed, problematic. They are highly sensitive to irrelevant factors and can be affected by such interventions as providing candy to students on the day of the evaluations. Students give better evaluations to teachers in classes where they receive higher grades. In one famous experiment an actor was hired to lecture on "the application of mathematics to human nature" and gave a lecture, peppered with anecdote and wit, but completely free of content. His instructional skills received excellent evaluations from audiences of psychiatrists, psychologists, social workers, mental health educators, and thirty-three educators and administrators enrolled in a graduate-level university educational philosophy course.[12]

Despite the numerous problems with student evaluations, it has proved extremely difficult to establish a causal link between giving high grades and receiving favorable evaluations. Valen Johnson's excellent book, *Grade Inflation: A Crisis in College Education*,[13] documents these difficulties, and even he ends up being unable to pin down the direction of causation, though he does successfully impugn the use of student evaluations of teaching for any public purpose.

Before moving on to the question of grade variation, I want to make a further clarification of the idea of grade inflation. Mansfield does, in fact, consider one of the possible confounding factors, and what he says about it is quite revealing. He writes that:

> Some say that Harvard students are better these days and deserve higher grades. But if they are in some measures better, the proper response is to raise our standards and demand more of our students. Cars are better-made now than they used to be. So when buying a car would you be satisfied with one that was as good as they used to be?[14]

Mansfield, of course, does not think that the students are better these days. But his response is interesting because it specifies circumstances in which he thinks that grade *deflation* ought to occur. Given any starting point, he thinks that we should make high grades harder to attain as the students improve. He does not, though, seem to think that standards should be lowered as the students get worse. It is hard to come up with a rationale for this asymmetry in line with the purposes of grading I have offered. A single institution's unilaterally raising standards for the allocation of grades does nothing to help grades inform prospective employers or students themselves. It might help with the pedagogical function of grades, but it might not; it will all depend on the students themselves and the culture they inhabit. It may have the effect of enabling faculty to feel that they are tough and hardheaded, and that is certainly valuable for *them*. But that benefit does not justify the effort it would require to achieve it.

GRADE VARIATION

Most of the public debate about grading has focused on inflation, about which the evidence is, as we have seen, less than conclusive. But the most sophisticated book-length analysis of grading practices, despite its name, focuses hardly at all on inflation, and much more closely on grade variation. Grade variation occurs when students within an institution receive different grades for similar-quality work within a single institution. At its most stark, this could occur within a single course—Celia receives an A for the same quality of work for which Betty receives a B. Considering its occurrence within a single course enables us to see the first, and most obvious reason, that something is wrong with grade variation: it seems to be straightforwardly unfair.

Before looking at other possible instances of grade variation, it is worth noting that grade variation within a course is exactly what you

might expect to occur if grading is used for the pedagogical purpose I mentioned in the first section. The professor might have judged that Celia responds better to getting slightly higher grades than her work, strictly speaking, merits, and that Betty, in contrast, will respond better to having the bar set a bit higher for her. The consequent unfairness is justified insofar as the pedagogical aim is to prompt the students to learn rather than to distribute amongst them a scarce resource. Since I think that prompting learning is more important than distributing grades, some unfairness is entirely tolerable, if it is the consequence of using grade assignments effectively to the end of inducing the effort of the student.

Unfairness is not the only reason for objecting to grade variation. If grade variation occurs across departments within an institution, then that might distort the choices students make concerning which courses to take. Suppose that English systematically gives higher grades for work of the same quality that receives lower grades in mathematics. Insofar as a student is concerned with protecting her GPA a student therefore has a reason to take English rather than mathematics courses. Such grade variation is deceptive and might lead students to make suboptimal choices for themselves.

How bad is grade variation, if it truly occurs? The concern about distortion of choice has to be tempered a little. What, precisely, are universities trying to do for their students? Consider two different answers. One answer has it that the university should be trying to enable the student to flourish intellectually. The other charges the university with enhancing the student's human capital to provide her with more economic opportunities. Call the first the *liberal* and the second the *vocational* answer. On the vocational answer, the distortion putatively caused by grade variation is clearly bad. The total sum of human capital is diminished because students end up failing to develop their more productive talents. But on the liberal answer, it is not so clear, for two reasons. First, there are numerous non-grade-related incentives to take the courses that are supposedly graded less generally. Even with lower grades, mathematics, computer science, engineering, and science graduates can earn more in the economy than their higher-graded contemporaries in the humanities. Grades are not the only currency to which students are sensitive. Higher grades in the humanities might simply compensate for the higher salaries attached to mathematics and science degrees. They might, in other

words, help correct for a different incentive students have to disregard their own intellectual flourishing. The second reason the liberal answer might be less troubled by the grade variation we are positing is that the university's responsibility for the students' choices is limited on this view. If students are moved by grade-enhancement considerations rather than by their own intellectual interests, then there is only so much that the university should invest in preventing them from doing so. As long as the university is not systematically or recklessly misleading the students about their own talents, and as long as it is providing courses that are intellectually valuable, it is not doing them a wrong. The GPA-obsessed student is wrongheaded, and that is *her* fault, not the university's. The university must facilitate, but need not (and probably cannot) force, the intellectual flourishing of the student.

Note that the mechanism prompting the choice matters a good deal on the liberal answer. I have assumed that the student in question is self-consciously GPA concerned. But there is a worse case, in which the liberal answer does impugn the university for the choices the student makes. This is where the student is quite unaware of grade variation and so takes the better grades in, say, English to be a signal that she is more talented in that field than in mathematics, and, wanting to develop her most fertile talents, she opts to major in English. If grades are portrayed as reliable guides to one's level of talent, then in this case the institution is complicit in deceiving the student about her own talents and hence is placing a barrier in the way of her intellectual flourishing.

I shall not argue for it here, but I would give very little weight to the vocational take on the university's mission, and much more to the liberal take, so I shall proceed on the unargued-for assumption that the liberal take is better.[15] On the liberal take, the distortion of choice that grade variation might produce is not of great concern. If variation diminishes the total supply of human capital, then that is not a problem; it is only a problem if it undermines the student's pursuit of her own intellectual development. But fairness might still be a problem on the liberal account; why should a student get "less" of something that matters to her, and the consequent lack of recognition and diminished self-esteem just because she has taken a course in a different department, or from a different professor?

But is grade variation a genuine phenomenon? Those who claim there is grade inflation also typically claim that there is grade variation

between departments. The overtime studies that show an increase in the median grade show that the increase is greater in humanities than in the social sciences, and greater in the social sciences (other than economics) than in the natural sciences.[16] The conclusion is that the sciences and social sciences grade less generously than the humanities, so there is, indeed, grade variation.

However, that conclusion cannot be reached so easily. The studies claiming to find variability do compare the performance of individual students in different kinds of class. But they lack a discipline-neutral standard of excellence. Considering some specific examples will illuminate. Which is more excellent, a fine piece of literary criticism or an elegant proof of a relatively trivial theorem? A high-quality painting or an essay presenting an objection to Descartes's method of doubt which, though familiar in the literature, the student arrived at by herself? An essay demonstrating that a large-scale study has failed to control for a crucial variable or a fine piece of blank verse? The scholars studying grade variation lack answers to these questions, and therefore they have no idea whether the achievements in the different disciplines merit the same, or different, median grades.

Doesn't the difference in the rate of increase of grades between the disciplines in itself provide evidence that there is variation? Or at least, doesn't it show that if there is no variation now it would be quite remarkable because, if there is no variation now, given the different rates of change, there has always been variation in the past? No. In the absence of some discipline-neutral standard we simply cannot make a judgment. Even if we did have some discipline-neutral standard, that might not be enough. The best predictor of one's achievement in a class is one's prior achievement in that class's subject matter. So if students are better prepared coming out of high school for classes in the humanities than in the natural sciences (and intermediately well prepared in the social sciences which, though they do not take them in school, draw on their preparation in both the humanities and the sciences) we would expect them to achieve more in the humanities. High school mathematics and science are the main shortage areas, and the shortages are worse now (relative to the humanities) than they have been in the past. This is usually taken as evidence that teaching in science and mathematics is worse than teaching in English and history, and worse, relatively, than in the past, hence the frequent calls for school districts to be allowed to provide incentive payments

in shortage areas such as science and mathematics. If so, then students are presumably less well prepared in the sciences than in the humanities, and that would provide a reason to expect their achievement in college science courses to be lower. Certainly, in the United Kingdom, for example, there is a widespread view within university science and mathematics departments that secondary school preparation has worsened significantly over the past decade or so, whereas there is no similar perception of crisis in university humanities departments.[17]

A SIMPLE REFORM

Suppose that both grade inflation and grade variation are genuine phenomena, and that the costs they impose are sufficient to justify some sort of reform. What reform should we adopt? Valen Johnson proposes an extremely complex arrangement that reconfigures GPAs so that they reflect the performance of the students relative to their peers in other classes.[18] Crudely put, professors would assign the grades they wanted to, but the grade would be deflated to reflect the general grading behavior of that professor. An A from a professor who gives many As would be worth less than an A from someone who gives out few. Johnson's proposal has a number of problems; it does not deal well with the problem discussed earlier, that we do not have discipline-neutral standards of achievement, and it would introduce the perverse incentive for high-achieving students to avoid high-grading professors even if those high-grading professors were the most excellent teachers. But any proposal will have drawbacks, and I do not think these are decisive, absent discussion of alternatives.

More telling is that there would be a great deal of faculty resistance to this, and, perhaps more significantly, a great deal of student resistance. The reform threatens the ability of students to manipulate their GPAs through choosing classes, and puts them more openly in direct competition with each other. Johnson's proposal also has the drawback of being transparently opaque; it is hard to imagine it being adopted as a result of university-wide discussions, because it is hard to imagine enough people being able to understand the complex details of the proposal for them to be immune to the attempts of ill-willed activists to undermine support for it. I have studied Johnson's book and proposal in detail, and I like it partly because it appeals to my

taste for complexity. But I would not have the confidence or patience to participate in a campus-wide debate on it.

I want to offer an alternative reform. European readers can switch off at this point, because what I shall propose reflects normal practice in secondary and higher education in Europe. It is based on the idea of dividing up the two roles of teaching and grading. The purpose of the teacher is to teach or, to put it in a more success-oriented way, to induce the student to learn. I am convinced that the teacher would be better positioned to achieve this end if someone else evaluated what the student had learned; or at least if someone else were the final arbiter of the grade the student received.

It is an interesting symptom of how different the educational cultures of the United States and Europe are that this proposal, and the view of the teacher's role underlying it, which is pervasive in Europe, can be described as being on the far end of radicalism in the United States. In their attempt to diagnose the causes of grade inflation, Henry Rosovsky and Matthew Hartley evince faculty hostility to the role of grading:

> Advocates of this opinion [that a higher grade might be given to motivate those who are anxious or poorly prepared] contend that students ought to be encouraged to learn and that grades can distort that process by motivating students to compete only for grades. . . . A *more radical view* holds that it is inappropriate for a professor to perform the assessment function because it violates the relationship that should exist between a faculty member and students engaged in the collaborative process of inquiry.[19]

This "more radical view" does not strike me as radical in the slightest. Anyone who has performed the role of both teacher and ultimate assessor for a student or set of students is aware that the two roles are in tension, and that the fact that one person embodies both roles inhibits that person's ability to perform both properly. Consider the role of a gymnastics coach. The coach works as hard as possible on behalf of her athlete, training her, trying to bring out her potential, pacing her, teaching her, and so on. Imagine if, in addition to being the coach, she were also the ultimate judge and, for that matter, the

person who set the rules (which is what the professor/grader is). Would she be tempted to bend the standards, or tilt the rules, in her charge's favor? Certainly she would. If she resisted that temptation that would show strength of character, but it would not show that it had not interfered with her carrying out the role. Now imagine that the gymnast *knows* that the coach sets the rules and is the final judge. Insofar as she wants to win, rather perform the very best that she can, she has an incentive to try to get the coach to tilt the rules, or skew the decisions, in her favor.

Having a single person embody both roles makes it harder for the teacher to perform her teaching role. This is because, like the gymnast, the student suddenly has an incentive to direct less of her effort to learning, and some, at least, of it to inducing the professor to make fewer demands of her. Even if she is entirely unsuccessful in her efforts, she has endured a loss, because she has put effort into trying that could more fruitfully be put into learning. Because students have limited knowledge of any given professor's tendencies, but do understand the incentives involved, they have reasons to try it on, even if the person they are trying to persuade is, in fact, unmovable.

Contrast this with the incentive structure when the ultimate arbiter of the grade is an external examiner to whom the student is anonymous. Now the professor and student have every incentive to cooperate to try and meet the external standards imposed. The student gains nothing by getting the professor to lower her standards, because even if she succeeds this will only result in less learning and a lower grade. Similarly, the professor has an incentive to teach the students well insofar as she cares about improving their grades.

So the proposal is to reform grading so that the teaching role and the ultimate assessment roles are divided. Different colleges and perhaps even different departments might pursue this in different ways. On the radical version of the proposal idea in any given course papers and exams would be graded by two faculty members who are not teaching the course, and every paper and exam would be graded by two people, with a third on hand to decide disagreements. On the moderate version of the proposal faculty would grade the work of the students they taught, but an external moderator would check the grades against an agreed standard that would bind all faculty in the department.

This proposal would do nothing to change the variation in grading practices *between* departments. I do not see how any proposal could

do that without solving the problem of cross-disciplinary standards. Nor would it do much about grade inflation. Departments, rather than individual professors, would have incentives to inflate their grades insofar as students pick majors to manipulate their GPAs. But it would do something to making grading practices within departments more uniform, and thus less unfair between students who take the same major and get different within-major GPAs as a result of the department's laissez faire approach to grading. It also would have the effect of inducing faculty within research institutions to spend more time discussing teaching and their standards and expectations for undergraduates, and it would prompt a more collaborative ethos concerning teaching. All these would be good things, but better would be the restoration of the relationship of teacher and student to its appropriate place: collaborators with the aim of maximizing the student's learning, rather than, in part, conspirators aiming to maximize the student's grade, on the one hand, or combatants tussling over it.

Consider, finally, an objection to the motivation for the proposal. In my discussions of institution-wide grade variation and multi-institution grade inflation, I have expressed a good deal of skepticism, and have, essentially, shifted the burden of proof back onto those who claim it does exist. This is because university-wide action is costly, and any that is recommended had better have a significant payoff; unless grade variation is a real problem, I see no reason to engage in university-wide action. But is there any evidence of grade variation within or across courses within a department? All of my evidence is anecdotal, gathered from talking to students and professors, many of whom are convinced (as I am) that it exists, but without any statistical evidence, let alone the sophisticated statistical and conceptual analysis I am demanding from those who claim system-wide problems. Why should the professor, or the department, not shift the burden of proof back onto me?

Think about in-course grade variation and the responsibility of the professor. I think the burden is on me to ensure that there is no grade variation within the course (except that marginal variation that serves individualized pedagogical purposes). This is because everything *other than* the grade is constant in the classroom. The students attend the same lectures, read and learn the same material, and receive similar-quality feedback on their written work and classroom participation. Whereas the same grades in Math 401 and English 401 send different signals to students and to third parties, the same grades in

Philosophy 341 taken in the same semester in the same department send the same signal. Just as the professor has a burden to ensure consistency in his or her own grading behavior, so does a department, in a way that a university does not have a burden to ensure consistency across departments, because so much else varies, and therefore so many other goods are at stake.

NOTES

1. Harvey Mansfield, "Grade Inflation: It's Time to Face the Facts," *Chronicle of Higher Education* (April 6, 2001): B24.

2. Again, reflecting on my own experience, as is normal in the United Kingdom (UK) I studied just one subject, philosophy, and assumed rightly and reasonably, that intellectual development in any particular area of Philosophy would have transferable benefits in other areas.

3. Mansfield, "Grade Inflation," B24: "Grade inflation has resulted from the emphasis in American education on the notion of self-esteem. According to that therapeutic notion, the purpose of education is to make students feel capable and empowered. So to grade them, or to grade them strictly, is cruel and dehumanizing."

4. In Alfie Kohn, "The Dangerous Myth of Grade Inflation," *Chronicle of Higher Education* (November 8, 2002), the author cites evidence from experimental psychology that focusing on "extrinsic' motivators (like grades) can actually inhibit intrinsic motivation. While I do not share Kohn's general hostility to grading per se, it is worth noting how different personal interactions are from grades. Although personal interaction (praise, enthusiasm for the student's idea, criticism, etc.) is in some sense extrinsic, it is hard for it to form an extrinsic motivation. The student is unlikely (in contemporary American culture) to seek the fulsome praise of the professor as her reward, and this can make it that much more rewarding and encouraging when it is forthcoming. Conversely, comments like, "This paper seems to have been written the night before without much thought—you can do much better, if you are willing to put in some thought and work," reveal much more thought and concern for the student than just giving him a very low or failing grade and (in my experience) yield better results.

5. The student in question would have found the very idea of reserving a grade for him absurd, laughable, arrogant, and vain.

6. On reflection I realize that I have not yet come across a student whose work is extremely good *and* who is sufficiently grade-obsessed that adding a reserved high grade would motivate or reward him or her at all in the presence of any of the alternatives I have mentioned.

7. Kohn, "The Dangerous Myth of Grade Inflation."

8. Ibid.

9. "Grade Inflation at American Colleges and Universities" (August 16, 2005), http:// *www.gradeinflation.com*. The universities involved were Alabama, California-Irvine, Carleton, Duke, Florida, Georgia Tech, Hampden-Sydney, Harvard, Harvey Mudd, Nebraska-Kearney, North Carolina-Chapel Hill, North Carolina-Greensboro, Northern Michigan, Pomona, Princeton, Purdue, Texas, University of Washington, Utah, Wheaton (Illinois), Winthrop, and Wisconsin-La Crosse.

10. I owe this observation to Daniel Drezner, who most assuredly should not be held responsible for the use to which I am putting it.

11. Of course, all of this gain could have occurred in pre-college education, and none of it in college education, but even if it had, college students would start at a higher level of achievement, so that holding constant the value added by their college education their achievement in college should be higher than before.

12. Reported in Valen Johnson, *Grade Inflation: A Crisis in College Education* (New York: Springer-Verlag, 2003), 134–36. The original paper is D. H. Naftulin, J. E. Ware, and F. A. Donnelly, "The Doctor Fox Lecture: A Paradigm of Educational Seduction," *Journal of Medical Education* 48 (1973): 630–35.

13. Johnson, *Grade Inflation.*

14. Mansfield, "Grade Inflation," B24.

15. See Chapters 2 and 3 of Harry Brighouse, *On Education* (London: Routledge, 2005), for arguments supporting the liberal take on the purpose of compulsory schooling.

16. See Johnson, *Grade Inflation,* 205.

17. See, for a good summary of the problem with respect to mathematics, the final report into post-14 Mathematics Education, Adrian Smith, "Making Mathematics Count," available on the UK government's Department for Education and Skills Web site at http://www.dfes.gov.uk/mathsinquiry/Maths_Final.pdf (accessed February 2004); see especially Chapter 2.

18. Johnson, *Grade Inflation*, 209–24.

19. Henry Rosovsky and Matthew Hartley, "Evaluation and the Academy: Are We Doing the Right Thing?: Grade Inflation and Letters of Recommendation," Occasional Paper (Cambridge, MA: American Academy of Arts and Sciences, 2002), 3.

This page intentionally left blank.

CHAPTER FIVE

From Here to Equality: Grading Policies for Egalitarians

FRANCIS K. SCHRAG

This chapter attempts to sponsor the marriage of an abstract, theoretical issue to a mundane, educational one. The abstract issue is the appropriate basis for social rewards; the practical one the appropriate basis for allocating grades in school and college.[1] This may seem to be an odd couple but it is a potentially fecund one, or so I shall try to show.

After considering objections to the proposed marriage in the first section I identify a powerful conception of egalitarianism in the second. In the third section I identify an important requisite for the emergence of a more egalitarian society. I discuss in the fourth section the ethics of grading, and in the final section I identify a variety of grading policies consistent with an egalitarian agenda.

Before beginning, however, what do I mean by grading? Let us distinguish providing feedback on a student's output, be it a dance or an essay, from the assignment of a letter or number to indicate the level of mastery the student has attained at the end of a course of study. I take it that the former, being indispensable to the student's development, is uncontroversial. I shall reserve the term *grading* for the latter practice.

WHY LINK JUSTICE TO GRADING?

Some may contend that egalitarians should simply work to abolish grading altogether, because it will disappear in an egalitarian society, one governed by the Marxian slogan "From each according to his ability, to each according to his needs." Is this really so? Even such a society will wish to facilitate a good match between what people are

good at and the services they provide to others. That being so, it will be in the interests of young people themselves to form a judgment of their strengths relative to those of their peers before deciding to invest years, sometimes decades, in preparing for an occupation.

In some domains, where performance is public and talent easily recognized, such as chess, tennis, or singing, grading will be unnecessary. Not all domains permit such ready gauging of capabilities, however. At the point at which students begin to seriously think about careers, especially careers that require the ability to master academic subjects, a refusal to assign grades to students serves neither the students themselves nor the society—even if it is an egalitarian society that severs the connection between the roles individuals play and the income they earn. Of course, the matching can be done through a test or series of tests at the end of high school, but there is little reason to think that it would be wise for young people or society to base career decisions on a single test or even a battery of tests.

No doubt many egalitarians recognize the need for grading even in an egalitarian society and take the opposite view—that their espousal of egalitarianism need have *no* implications for their grading of students.[2] Egalitarianism, they may argue, need not imply the elimination of competition and selection for occupations. Such competition, they will claim, makes an efficient allocation of people to occupations more likely. Of course, the argument continues, society will want to provide some level of equality of opportunity, for talent can be found in all social precincts. But even if the children of the more favored social classes are more likely to prevail in the competition (as is likely), a sufficiently progressive tax can, the argument concludes, redistribute income so that these advantages work to the benefit of all.

Moreover, egalitarians may argue that they are concerned about the distribution of social *rewards*, and grades are not rewards in this sense but simply indicators of the quality of a student's product or performance. It is possible to think of grades as mere indicators in this sense, but to see them that way *only* is to miss the crucial role they play in *determining* the distribution of vital social benefits: wealth, power, and prestige. Without high grades, a high school student has little chance of admission to a selective college, and a college student has little chance of admission to a selective professional school. The kinds of jobs or professional schools open to graduates of Princeton

or Berkeley are simply not the same as those open to graduates of Seton Hall or California State University.

But it is the first contention that I wish to challenge, the contention that a meritocratic reward and selection system can be in place while children are at school, with egalitarian policies kicking in afterward when students join the workforce. The problem is that meritocratic grading policies cannot support egalitarian dispositions. Since understanding this point depends on an understanding of egalitarianism itself, I turn to a brief exposition of that concept.

EGALITARIANISM

Egalitarian principles have been the subject of vigorous contemporary controversy since the publication of John Rawls's *Theory of Justice* in 1971.[3] It is not necessary to summarize the enormous literature or to formulate my own conception. Instead I shall identify rival concepts of egalitarianism, each of which has been defended in the recent philosophical literature.

One prevalent view, which following Elizabeth Anderson I will label "equality of fortune" or "luck egalitarianism," is identified with, among others, Rawls, Ronald Dwordkin, Richard Arneson, John Roemer and G. A. Cohen.[4] In "Where the Action Is: On the Site of Distributive Justice," Cohen states the nub of this position:

> My root belief is that there is injustice in distribution when inequality of goods reflects not such things as differences in arduousness of different people's labors, or people's different preferences and choices with respect to income and leisure, but myriad forms of lucky and unlucky circumstance.[45]

Elsewhere Cohen articulates the key idea boldly and succinctly:

> In my view, a large part of the fundamental egalitarian aim is to extinguish the influence of brute luck on distribution. Brute luck is an enemy of just equality, and, since effects of genuine choice contrast with brute luck, genuine choice excuses otherwise unacceptable inequalities.[6]

Although this summarizes the core of the view, it is important to recognize that luck egalitarianism represents not one but a family of views. Luck egalitarians differ on the degree to which rewards ought to accrue to talent. John Roemer takes a much more austere view here than does Rawls. Despite such differences, luck egalitarians maintain that in the allocation of social rewards, those factors that are within a person's control ought to count far more than those that are not, and they agree that talent is largely a result of what Rawls calls the "natural lottery."

To make luck egalitarianism more vivid consider two college students from the same high school, Jack at Seton Hall University and Jason at Princeton. Jason can in all likelihood anticipate far greater rewards—income, power, and prestige—than Jack. Is this fair?

Luck egalitarians say that if from middle school on, Jason made scholastic excellence his primary commitment, a commitment reflected in the effort he put into his studies, whereas Jack chose to spend a good deal of time lounging around listening to CDs, then the disproportionate rewards are fair. Suppose, on the other hand, that effort and commitment by the two students were about equal, but Jason, who let us assume scored somewhat higher on the SAT and whose father is an alumnus, was admitted to Princeton while Jack was not. Jason's prospects exceed Jack's by a substantial margin. (Citing Virginia Valian's *Why So Slow?,* Natalie Angier writes, "A degree from a high-prestige school contributes $11,500 to a man's income."[7]) Is this fair? Defenders of a meritocratic reward system could point to Jason's greater academic aptitude (as measured by his SAT score), but this point is not decisive according to the luck egalitarian. Whether such aptitude is a measure of native intelligence or an index of cultural capital acquired in his family, the results are still due *to brute luck as far as Jason is concerned.*

When Jason's economic prospects substantially exceed Jack's for reasons that have nothing to do with either student's efforts or choices and everything to do with their circumstances (which we may take to include their genetic endowment), this strikes luck egalitarians as unjust. The distinction between people's choices, including choices to expend effort, on the one hand, and their circumstances, on the other—the key distinction emphasized by luck egalitarians—is one whose moral significance is easy to appreciate; of course, luck egalitarians recognize the difficulty in practice of determining what proportion of a person's accomplishment is due to brute luck.

I myself find this conception convincing, but I recognize the powerful critique of luck egalitarianism offered by Anderson, who argues that egalitarians ought to pursue a different kind of equality, which she calls "democratic egalitarianism." Others with whom I would associate her view include Nancy Fraser, Michael Walser, David Miller, and John Dewey.

> The proper negative aim of egalitarian justice is not to eliminate the impact of brute luck from human affairs, but to end oppression, which by definition is socially imposed. Its proper positive aim is not to ensure that everyone gets what they morally deserve, but to create a community in which people stand in relations of equality to others. . . . In seeking the construction of a community of equals, democratic equality integrates principles of distribution with the expressive demands of equal respect.[78]

In articulating the notions of a community of equals that democratic equality demands, Anderson writes as follows:

> The democratic egalitarian is not insensitive to inequalities of income or resources but "would urge a less demanding form of reciprocity. . . . The degree of acceptable income inequality would depend on how easy it was to convert income into status inequality—differences in the social bases of self-respect, influence over elections, and the like."[9]

Although Anderson's concerns are legitimate, this last sentence weakens her case for democratic equality in our own society. Nothing is clearer than the ease with which inequalities of income are, in fact, convertible. As Adam Swift notes:

> However we may try to block exchanges between goods or dimensions of advantage, however successful we may be in blocking particular channels, there will always and inevitably be modes of conversion that we cannot prevent. If advantage is Protean, then the task of fettering it is Sysyphean.[10]

How can we prevent the more privileged, like Jason's father, from converting such income into more political power and social status for themselves and their children? Even if Jack's diligence had equaled

Jason's (case 2), Jack's chances of becoming a member of the national elite are limited by the mere fact of his attending Seton Hall, while Jason attends Princeton. The problem is not that Jack's academic program will be less rigorous than Jason's (though this may be true), but that Jack will not acquire the kind of cultural capital and establish the sorts of social connections that will enhance his family's political influence and social status, while Jason will. The democratic egalitarian is hard put to block this kind of conversion of one form of advantage into another.

While proponents of equality of fortune and democratic equality differ markedly in some of their conclusions, note that their views overlap considerably. Both groups denounce not only formal barriers to equal citizenship or participation but barriers that result from unequally distributed resources. Both groups admit that extreme inequalities of wealth and income pose a danger to the abilities of all citizens to lead good lives. Most luck egalitarians concede that some measure of *in*equality may be necessary if their efforts to protect the worst off are to succeed. Finally, it is noteworthy that Cohen, a luck egalitarian, considers community a coequal value, although he defines community in his own socialist terms as "the anti-market principle according to which I serve you not because of what I can get out of doing so but because you need my service."[11]

The values of community and of equality of fortune, far from being antagonistic, are to a considerable extent mutually supportive. The sense of community is likely to hold down differences between rich and poor. Reducing the effects of brute luck on the distribution of wealth and income is likely to sustain the sense of community. Although the idea sounds utopian, it is well to remember that an appreciable measure of *both* kinds of equality have been realized on the Israel kibbutz during the early years of statehood. Even in such a community, however, there must be ways for students and for the community to find out what occupations individuals are best suited for, so even here grading is not eliminated though it may (and should in my view) be postponed until students need to start thinking about future careers.

FROM HERE TO EQUALITY

One difficulty in moving to a more egalitarian society derives from the fact that no matter how convincing an argument for egalitarianism

can be mounted, social transformations depend on more than philosophical argument. To take root in a population, new norms need to be instituted and experienced. Of course, the institution of new norms depends on argument, but should the argument not connect to actions or proposals that those sympathetic to egalitarianism can actually take? Egalitarians cannot expect that a society that consistently sends one message to its youth throughout their schooling—expect recognition commensurate with your level of accomplishment compared to that of other students—will easily turn around and say to its adult members, don't forget you live among equal citizens; even the least productive of you deserves our respect; to say nothing of, remember, you are responsible for your efforts and choices only, not for your accomplishments. Expect rewards commensurate with those efforts and choices. Neither a democratic nor a luck egalitarian would dissent from Cohen's assertion that "a society that is just . . . requires not simply just coercive rules but also an *ethos* of justice that informs individual choices."[12]

Consider for a moment the revolution in the social expectations of and for women as seen in the fact that today's fathers and mothers have *virtually* the same expectations for their *daughter*s as for their sons. In *Destined for Equality: The Inevitable Rise of Women's Status*, Robert Max Jackson shows the role that coeducation played in this transformation. "Schools, particularly coeducational schools, put the idea of male intellectual superiority to the test. It failed. In schools, girls and boys' experiences vividly belied beliefs in unequal intellectual potentials."[13]

As male claims to female intellectual inferiority "gradually became more inconsistent with people's *experience,* . . . they became more vulnerable to challenge."[14] I will not speculate about how a more egalitarian society might actually evolve here in the United States. One thing is clear to me, however: for such a society to emerge, our propensity to justify inequalities of status, power, and condition in terms of inequalities of accomplishment must be weakened. For people to grow up disposed to more egalitarian structures, they must *experience* them during their socialization. That is why I wish to examine the implications of egalitarian justice for grading students in school and college.

GRADING CRITERIA

Why grading? For two reasons*:* First, because for most of our children, school grades are one manifestation of the way in which society

regards them, and because grading practices convey norms that students are expected to internalize. As sociologist Robert Dreeben argued a generation ago, one of the school's principal roles is to convey through its own normative structure the normative structure of society beyond the family.[15] Second, because chances are that you who are reading this participate in grading practices and hence play a role in the perpetuation or subversion of society's normative structure, a role that is worth bringing to conscious focus, even if you do not share my egalitarian sentiments. If you do share these sentiments, you ought to reflect those in your own grading practices. Of course, an egalitarian agenda in education will go beyond grading, but my focus here is on grading as one central practice that all teachers participate in.

Before discussing grading policies designed to foster a more egalitarian ethos, however, we need to identify the desiderata of any ethical grading policy. I identify four: (1) *Grades should not convey deceptive information to those who receive them.* Grades typically send signals to a variety of audiences in addition to the students themselves: prospective employers, college or graduate schools' admissions committees, and parents are the most important. Most readers of transcripts are likely to interpret grades and transcripts in fairly predictable ways. When, for example, a college admissions committee sees high school transcripts recording Jack as having earned a B in world history and Jason an A in the same class, committee members will infer that the quality of Jason's work was superior to Jack's. If they act on interpretations that turn out to be mistaken, then they will likely be committing injustices.

(2) *The educational climate created by the policy should not gratuitously impede the learning of any student*. If the educational experience is designed to foster students' growth in the fields they study, then a grading policy that raises an unnecessary barrier to such growth would not meet this test. Let me give two examples: (1) A course in which the bar for a passing grade is raised very high may depress the motivation of weaker students. If a student sees a very small chance that her effort will be rewarded with a passing grade, then she may give up and so learn less. (2) When grading is a zero-sum game, as in grading on a curve, a stronger student has a disincentive to help a weaker student master the material, thereby limiting the latter's opportunity to learn from her peers. These strike me as defects in a grading scheme from the ethical point of view. (Grading on a curve is objectionable on other grounds, which I touch on later.)

(3) *The grading policy should not subvert ethical dispositions.* This is closely related to the previous point: a grading scheme that penalizes sharing and mutual assistance or one that creates excessive pressures to cheat or lie—for example, by not allowing makeups for illness or family emergencies—fails to meet this requirement.

(4) *The grading policy is fair to the students being graded, and it is perceived to be so.* Although both luck and democratic egalitarians would subscribe to this criterion, they would interpret and implement it rather differently, so let us consider each in turn.

Since a democratic egalitarian like Anderson has no objection to income inequalities per se, she would not necessarily have a problem with grading students according to the quality of work they produce, though she would try to avoid having these judgments undermine the needed sense of respect for all citizens in a democratic community. Although a democratic egalitarian will surely regret the way grading in school often reinforces *existing* status differences, she will not consider assigning grades on the quality of student work to be per se unfair. Not so the luck egalitarian, on whom we focus the remainder of the discussion.

What is a fair grading scheme for the luck egalitarian? To whom (or what) should the grader be fair? There are at least two answers: to the work being graded and to the student whose work it is. Consider them in order. Teachers and professors see one of their jobs as cultivating in students the ability not only to learn the facts and skills connected to a field of study but also an ability to understand and judge what gives work in that field value. When teachers grade students, especially at more advanced levels, they are sending messages about the quality of student work in the hope that the students will internalize their judgments. Grades are not necessary to provide meaningful feedback on the strengths and weaknesses of student work, but where they are employed at the end of a course, if only to give students a sense of their overall *level* of accomplishment, instructors will normally and appropriately feel dishonest if they give the seal of approval to inadequate performance.

For the luck egalitarian, fairness to *students* is not the same thing as fairness to the work. Taking Cohen's distinction as cited earlier, it is unfair to reward or penalize students on the basis of "brute luck," which includes the gifts with which they are born, those that happened to be nurtured in the milieu in which they grow up, and those resulting

from the chance exposure to the more or less adequate teachers to which they have been assigned. Thus if a student who has produced a paper with enormous effort receives a higher grade than a student who has produced a paper with little effort, then is this just *even when the second paper is higher in quality than the first?*

Many will say this is unfair, but the luck egalitarian will counter that this is because they conflate grading the paper with grading the student. Professors should not mislead students about the quality of their work; students, recall, need to form realistic estimates of their strengths and limitations, *and* they need to know what is and what is not quality work and why—but that is no reason the grade for the work and the student's grade need to be one and the same. Indeed, according to the luck egalitarian, fairness demands that the two be separated.

GRADING POLICIES FOR THE LUCK EGALITARIAN

Alas, effort-sensitive grading is bedeviled by severe difficulties, as an article on effort-based distribution principles by Julian Lamont demonstrates.[16] Lamont begins by clarifying the relationship between effort and productivity, pointing out that even those who favor rewarding effort over productivity do not want to reward any old effort—doing push-ups, for example—only effort that leads or could be expected to lead to higher productivity in the area being evaluated (in our context, for instance, becoming a better all-around learner or a more accomplished student of world history). As Lamont shows, rewarding productivity and rewarding effort are not so much alternatives as distinct locations along a single spectrum. One of the attractive features of rewarding effort is that it typically—though not always—*enhances* productivity. Recall, furthermore, that both the student and the larger society have reason to care that students train for and enter occupations that will be reasonably well matched to their capabilities. The luck egalitarian instructor will not, therefore, want to reward *only* effort, but she will certainly want to *permit effort to count in determining students' grades*.

The principal difficulty facing effort-sensitive reward policies is that it is impossible to identify the proportion of a product or performance that is the result of effort alone. All products or performances are compounds of effort and trained capacity, with capacity resulting

from brute luck and choices. Capacity itself may be expanded or contracted as the result of effort. To add to the complexity, effort may be influenced by brute luck, for example, because of a chemical imbalance of unknown origin, Joe may simply have trouble concentrating; Jason may have the time available to expend effort on schoolwork, while Jack's parents insist that he work after school. As a practical matter, therefore, it is virtually impossible to identify the contributions of effort and choices to a performance or product.

Attempts to resolve this difficulty raise what Lamont, following economists, calls the problem of "moral hazard." This is best illustrated by a concrete example. Since a student's product, say a paper, does not directly display the amount of effort that went into it, some proxy for effort will be needed. Suppose the professor, on the presumption that it takes more effort to write a longer than to write a shorter paper, announces that the number of pages will be a proxy for effort. Now students have an incentive to write more *pages* rather than fewer, but not necessarily to exert more *effort.* Indeed, Jack may write lots of pages while exerting less effort than Jason who writes fewer pages. The moral hazard is the risk of eliciting behavior that brings about the proxy rather that action that achieves the desired goal.

An additional difficulty facing use of the effort criterion derives from the requirement of sending honest messages to third parties. If most readers interpret grades as measures of quality and the luck egalitarian instructor intends to send a message that takes the student's effort into account, then the communication will be deceptive. I shall propose a way to mitigate this problem later.

Some may read the previous paragraphs as a *reductio ad absurdum* argument against any policy that tries to take effort into account. This conclusion is, I believe, premature, but I do not want to underestimate the difficulties of such a policy. The luck egalitarian faces two severe challenges: the first is to find a way of assessing effort that does not create a moral hazard; the second is to find a way of incorporating effort into the grade without sending deceptive messages about the quality of student performance to third parties.

I believe there is a way of incorporating effort into a student's grade that *reduces* (though it does not entirely eliminate) the moral hazard problem. This consists in permitting students to earn extra credit by choosing to expend additional effort on work that meets some minimal level of quality.

The following three actual illustrations are taken from college courses. The first, a large-lecture sociology course has both written assignments and multiple-choice exams. The professor, Erik Olin Wright, notes that the course contains extra readings on most topics, and that each exam will contain a special section asking questions on these readings. According to the syllabus,

> Grading on the exams will be done in the following manner: The scores required to get different grades will be determined strictly on the basis of the questions on the required readings. Correct answers to the extra-credit questions will then be added to your score for the required questions to determine your grade on the test. To prevent people from simply guessing answers in the extra-credit section of the test . . . incorrect answers will be subtracted from correct answers in the extra-credit section to determine the extra-credit score.[17]

The second illustration is furnished by my colleague Michael Apple. Apple provides students with an initial grade on their written work but invites them to rewrite their papers in response to the detailed comments and suggestions for improvement that he provides. The initial grade is then revised. Professor Apple promises to reread and regrade student papers as often as the students wish to rewrite them.[18] The third illustration comes from a course in educational ethics that I co-teach with Daniel Pekarsky. Among other requirements, students are given the option of writing a term paper on an approved topic of their choice or a take-home final. A student wishing to demonstrate extra effort may do both. The grading policy states that a student may raise his or her grade on the paper or the final by doing the other as well, provided the additional work is of at least B quality. Of the three examples, I think providing opportunities for revision of required work to raise a grade is soundest, because it focuses the effort in such a way as to enhance the quality and not just the amount of work produced by the student, and it minimizes, though it does not entirely eliminate, the moral hazard problem. Suppose, though, that the student's efforts to revise a paper, while evident, do not succeed in actually improving it, or the paper is better in one dimension but worse in another. What then? The student might be given further guidance and opportunity to improve the work, but this would be to reward effort

only if it increased productivity. An effort-rewarding policy ought to reward effort, regardless of whether the effort succeeds. But this does not mean that the student's *work* ought to be judged to have higher quality than it actually has, or that the student should not be aware that his or her efforts have not enhanced the quality of the paper.

Note that even if these policies do come to grips with the problem of moral hazard, they are far from ideal: There is no way to know that a paper that was outstanding when it was first handed in was not the result of heroic effort, nor is there a way of determining the role brute luck played in determining the availability of time and energy to do extra work. Most important, the grades that do count effort convey misleading messages to third parties and are reprehensible on that account from an ethical point of view. This, I take it, is a large part of the concern that people have about "grade inflation."[19] My proposal to solve this problem unfortunately cannot be adopted by the individual professor acting alone. It requires institutional action, making the effort-based policy more transparent. For example, the transcript could indicate by an asterisk any grade earned by producing additional work not required by all students. This might be coupled with an understanding that additional effort could raise a grade no more than one level.[20] Until the passage of such a measure the egalitarian instructor must decide whether to give priority to supporting egalitarian justice or to avoiding deception. I do not think there is a clear answer here.

Neither the democratic nor the luck egalitarian conception supports a policy in which all students in a course *automatically* receive the same grade regardless of either effort or accomplishment, even if the grade is an A. Nor would grading on a curve based on test scores be consistent with either conception. The democratic egalitarian would object that grading on a curve forces a ranking among students that may have actually all succeeded about equally well or badly. It not only *recognizes* distinctions, it creates them. The luck egalitarian would object doubly: not only does grading on a curve fail to consider effort, but it makes a student's grade depend on the luck resulting from the mix of students that happened to enroll in the course that semester.

To conclude: Consider the problem of awarding grades to three student papers in a history class, those of Kirsten, Jane, and Bob. Kirsten's is brilliant, and she is clearly gone to great lengths to collect information and polish the prose. Jane has spent hours on her paper, rewritten it four times, and it is competent, but it is not imaginative.

Bob is an extremely talented student who rarely applies himself. His paper is brilliant, but it is clear that he is dashed it off in virtually no time. The position I have endorsed here, luck egalitarianism, holds that it is most just to rank Kirsten's performance, all things considered, above Jane's, and Jane's above Bob's. Such a ranking honors the fundamental distinction between choice and circumstance. It is a matter of decision or choice as to where effort should be placed, but there is no choice with regard to talent or imagination. A just grading practice will reflect this ranking. However, the challenge is to do so without sending deceptive messages about the quality of their papers—for instance, suggesting that Jane's *paper* is as good a paper as Bob's is—and without unduly creating moral hazards. Subject to these caveats, the best system will be one that honors Jane for her effort.[21]

NOTES

1. Although philosophical critiques of education in the late 1960s and 1970s would see the connection as quite natural, recent theorizing has tended to avoid consideration of mundane, practical matters such as grading. It is hard to point, for example, to a contemporary parallel to philosopher Robert Paul Wolff's *The Ideal of the University* (Boson: Beacon Press, 1969), which was addressed to a wide audience and forwarded practical proposals on grading and college admissions, proposals that grew out of Wolff's more scholarly analysis of the nature of the university.

2. Eric O. Wright, Personal communication.

3. John Rawls, *The Theory of Justice* (Cambridge, MA: Harvard University Press, 1971).

4. Elizabeth S. Anderson, "What Is the Point of Equality?," *Ethics* 109 (1999): 287–337. One excellent source for recent discussions of justice is Louis Pojman and Owen McLeod, *What Do We Deserve? A Reader on Justice and Desert* (New York: Oxford University Press, 1999).

5. G. A. Cohen, "Where the Action Is: On the Site of Distributive Justice," *Philosophy and Public Affairs* 26 (1997): 12.

6. G. A. Cohen, "On the Currency of Egalitarian Justice," *Ethics* 99 (July 1989): 931.

7. Natalie Angier, "Men, Women, Sex, and Darwin," *New York Times Magazine* (February 21, 1999): 50. Angier notes that according to Valian the same degree decreases a woman's salary by $2,400.

8. Anderson, "What Is the Point of Equality?," 288–89.

9. Ibid., 326.

10. Adam Swift, "The Sociology of Complex Equality," in *Pluralism, Justice, and Equality*, ed. David Miller and Michael Walzer (Oxford: Oxford University Press, 1995), 263.

11. Cohen, "Where the Action Is," 9.

12. Ibid., 10, emphasis in original.

13. Robert Max Jackson, *Destined for Equality: The Inevitable Rise of Women's Status* (Cambridge, MA: Harvard University Press, 1988), 155.

14. Ibid., 155–56.

15. Robert Dreeben, *On What Is Learned in School* (Reading, MA: Addison-Wesley, 1968).

16. Julian Lamont, "Problems for Effort-Based Distribution Principles," *Journal of Applied Philosophy* 12: 3 (1995): 215–29.

17. Erik O. Wright, *Syllabus for Sociology* 125, 3. Wright, personal communication, 1998, argues that though the reward structure is intended to fit his egalitarian beliefs, he prefers to see it as broadening rather than subverting the idea of merit.

18. Michael Apple, personal communication, 1998.

19. See the second section of Richard Kamber's chapter, "Grade Inflation as Epidemic" in this book.

20. Whether this would stigmatize the students earning a grade with an asterisk next to it is not clear to me; I view it as an empirical question.

21. Thanks to Harry Brighouse, Sandra Finn, Ken McGrew, Richard Merelman, Jennifer O'Day, Daniel Pekarsky, Amy Shuffelton, and my wife Sally for valuable feedback on earlier versions. Thanks, finally, to Nick Burbules for constructive suggestions for improving the version he originally read.

This page intentionally left blank.

CHAPTER SIX

Grade "Inflation" and the Professionalism of the Professoriate

MARY BIGGS

Let me tell you about three students:

The first student is Dustin, who took one of my General Education literature courses. While discussing the novel *Emma*, which the students had been assigned to read the week before, I asked him a question about Jane Austen's perspective. He looked around the room at his women classmates and said, "I don't know. Which one is Jane?" Dustin calmly acknowledged having read none of the assignments but protested his D at semester's end, pointing out that he had always been in class and saying, "You're supposed to get a B in Gen. Ed. if you come to class. A if you do all the work."

The second student said his name was Peter but to call him Ian, and he was registered as Andrew. After three or four weeks, he rarely attended class and submitted few assignments. When his transcript arrived, he telephoned me about his F, crying, "How could this *happen?* My parents are very upset, and my next lowest grade is C minus." "Then you must have attended your other classes regularly," I observed. "No," he wailed. "I didn't go to those either."

The third student is Ruth, who also cried—about a B. She had modest talents but worked very hard, and I believed her when she said it was the first B she had received in her life. "It's a *good* grade," I told her, and she insisted, "No. B is for *bad.*"

We all tell these stories—those of us who still give Ds, Fs . . . and Bs. And I am not being critical of Ruth, Dustin, or even Ian/Peter/Andrew. I am critical of the culture that shaped their assumptions.

I teach at The College of New Jersey, a highly selective, state-supported, mostly undergraduate institution that comprises four

professional schools in addition to the liberal arts and sciences. The most commonly awarded grade at my college is A. The second most commonly awarded is A–. The third is B+. And so forth, right on down the line with marvelous predictability. This has been true every semester at least since Spring 2001.[1] Put another way, in Autumn 2003, the modal grade awarded by thirty of our thirty-four departments was either A or A–, usually A. The modal grades of the four exceptions were B+ or B (two each).

In certain disciplines, grades are especially high. We supply New Jersey with many of its schoolteachers. In the same semester, Fall 2003, 74 percent of grades in Educational Administration/Secondary Education were A or A–, 70 percent of grades in Elementary/Early Childhood Education, and 62 percent in Special Education. In another critical professional program—Nursing—half of all grades were in the A range, 5 percent in the C range, and almost none were Ds or Fs. Our "normal curve" resembles a left riverbank.

It would seem that some disciplines, notably the natural and technological sciences, are inherently inhospitable to grade inflation.[2] So how could Physics award 31 percent As and A minuses, mathematics 30 percent, and Computer Science an astounding half, even more than the English Department's 48 percent? Now English is *my* field.

Let me tell you about three of my colleagues: Jack showed me the distribution of scores on his latest quiz. They formed a perfectly symmetrical curve: two As, five Bs, ten Cs, five Ds, and two Fs. We marveled. "It must have been a well-designed quiz," I complimented him. "Yes," he agreed, then sighed. "Now I have to jack up the grades. I hate this part."

My friend Joe explained why students never challenge the test grades he gives: "I have them choose the grade *they* think they deserve." "And does that stand as the *final* grade?" I asked. Incredulously. He smiled, shook his head, and revealed the special genius of his system: "Only if the grade is high enough. If *I* would grade the student higher, I change it."

And finally there is Jenny, about whom a student once told me: "She's so *nice*. She has two grades: A, good job, and A, nice try."

What do those three faculty—chosen almost at random to mention here—have in common, other than their lenient grading? All are tenured. One is a full professor, and a second does not intend to apply for promotion. *They have nothing to lose*. The names are changed, the details

are true, and yes, we all have stories like *this,* too, and many more that we could tell. Some may even be about *us.* For example, how could I have given Dustin a D?

Grade inflation exists. We have been hearing about it for decades,[3] of late most conspicuously through the self-publicized embarrassments of As, and modest proposals for reform, at first Harvard and then Princeton University. To advance the discussion of possible solutions, we must step over the red herring of denial, which is still, though increasingly rarely, thrust into our path.

What is more, grade inflation *is* occasion for concern, despite arguments that higher grades may somehow be deserved.[4] All evidence confirms that students are not better prepared or smarter than they were a few decades ago, and that they spend less, not more, time studying.[5] But if they *were* smarter, they still should be productively challenged. Any large human grouping includes a few people who perform at an excellent, or A, level; a few at the F level, who are seriously deficient by reason of ability or motivation; and a vast majority who lie somewhere between those extremes, the largest number being roughly midway. Average C. The heights of the levels of ability or achievement will vary by group, but multiple levels will always be disclosed in an environment structured with that intent. Having mentioned the C, I should add that even more striking than the proliferation of As and the elevation of medians and modes is the transformation of the lowest three grades. C is regarded by many students as a devastating insult, Ds are rare, and F has become shorthand for "seldom showed up" or "didn't do the work" rather than for "completed the course but not at an acceptable college level."

Still another red herring is the question of causation. In other works, Richard Kamber and I have taken a diverse historical approach to it. Suffice to say here that commentators tend to focus on one or a very few possible factors when the reality is far more complex and has decades- or even centuries-old roots. In the most wrongheaded and painful cases, they politicize the question by dragging forth scapegoats: members of some group historically underrepresented among college students. This scapegoating practice is very old, long predating grade inflation, for observers have always worried about the degradation of academic standards. The admission of farm boys was deplored in the late nineteenth century; then of immigrants, Catholics, women, Jews, veterans, Latinos; Negroes when that was the preferred term, then

blacks, then African Americans . . . almost every socially, economically, politically vulnerable group in this country that has shown ambition to achieve academically has been charged, at some time, by someone, with causing a compromise of academic standards and, more recently, with spurring grade inflation. Understandably, this has frightened and riled social liberals, and historical investigation does belie it. An unfortunate result is that critics of grade inflation have become identified, in some minds, with racism, classism, and ethnocentrism, or at least with the narrowest conservatism. As I have said, the debate over causes is nearly always superficial, is often politicized, sometimes gets ugly, and is ultimately beside the point I wish to argue here, which is *the cause of grade inflation is the faculty. We give inflated grades.*

There are reasons for this—which, in both oral and written discussion, often ramify as excuses and not infrequently deteriorate into whining. But the fact remains that *we are* the cause. When we award an A to a good but not outstanding student, we *could* give a B and perhaps stimulate that student to higher achievement, to the formulation of an inspiring goal. When we give an average performance a B, we *could* give it a C. And when we bestow a D upon a student who shows up and shows some effort but should not be certified as having acquired college-appropriate mastery, we *could* give that student an F.

When senior faculty use findings from poorly constructed student evaluation instruments in ways that affect the careers of their junior colleagues—affect them negatively *or* positively, and whether they are full time *or* adjuncts—they take the easy route to assessment of teaching: a route that, as Valen Johnson has so persuasively demonstrated and other, less impressive, studies also suggest, tends in a different and much less scenic direction for the rigorous grader than for the lenient one.[6]

But this is not to say that faculty at any level are justified in distorting grades to win votes from student evaluators. Some faculty insist they have no choice, and maybe *somewhere,* in *some* college or university, administrators are standing over instructors, holding their hands and forming the marks that their No. 2 pencils will make on their machine-readable grading sheets, or pushing their fingertips along the computer keys as they input their grades, but it has never happened to me or to anyone I know. More plausible is the possibility that somewhere, in some tuition-driven colleges, administrators are exerting pressure on faculty to at least pass the students, and perhaps even to grade so favorably that students will not become discouraged and drop out or

transfer to more charitable competitors. But it appears, in fact, that many less selective, less wealthy institutions also inflate grades less: less, that is, than the most selective, best-endowed colleges and universities, which scarcely need to worry that they will be unable to fill their classrooms and survive. (Have the easy A and guaranteed B become two more perquisites of class privilege in America?)

Still, it is possible that some faculty somewhere feel intense administrative pressure to inflate grades. It is certain, as Richard Kamber points out in another chapter in this book ("Understanding Grade Inflation"), that once exaggeratedly complimentary grading inserted an entering wedge into the system, it quickly created its own constituencies of people who found it easier, pleasanter, ego-enhancing, time-saving: utterly comfortable. These constituencies include faculty members themselves, of course, but what I have in mind now are students, tuition-paying parents, and, yes, administrators and trustees. And those constituencies *can* exert pressure that feels compelling.

So what? Professionals resist pressure to betray professional standards. That is the essence of professionalism. Its reverse defines *un*professional.

Let us return to Jack, Joe, and Jenny, the faculty anti-heroes of my true stories. All of them, like all of us, would bristle if their professional status was questioned. The professoriate shares with law, medicine, and the clergy, and perhaps only with them, an indisputable and a comparatively long-standing classification as a profession. And what *is* a profession? The literature of the history and sociology of American professions offers a range of answers, but writers tend to agree on certain elements. Some are obvious: professions found regional, national, and international associations and publications. Most formulate, interpret, and in some way enforce codes of ethics.

A professional possesses and applies specialized knowledge built on research and theory, typically obtained through a postsecondary program that has been accredited or certified by the profession itself, which also may set licensure requirements for individuals. The professional/client relationship is asymmetrical not only in expertise but also in autonomy and power. The ethical professional takes responsibility for not abusing that power and for justifying the client's trust. Perhaps most significantly, the professional is motivated by, in Burton Bledstein's words, "an ethic of service which teaches that dedication to a client's interest takes precedence over personal profit, when the two happen to conflict."[7] I want to clarify what the terms *client* and *personal profit*

mean when they are adapted first to any profession and then specifically to academics.

The immediate client, of course, is the person with a body to be healed or a soul to be saved, a case to be litigated or funds to be managed, or a mind to be shaped, filled, and disciplined for optimal participation in the larger society. But that larger society also is a client; it is, really, the ultimate client, dependent upon professions as practiced individual by individual: doctor by doctor, lawyer by lawyer, accountant by accountant, engineer by engineer, teacher by teacher, nurse by nurse. It is dependent upon them to develop and maintain the entire infrastructure that supports the kind of society in which we have decided we want to live—and for making it better. Violation of this ultimate client's trust can upend whole institutions, break our hearts both individually and collectively, as the numerous professional scandals of the past decade have made very clear. Arthur Andersen's perfidy makes investors wonder if any account sheet can be believed. The exposure of one doctor's profiteering through a drug company affiliation leaves thousands of sick and helpless people uneasy about prescribed treatments. But it takes just such catastrophic *public* events to produce wide-scale vocal reaction, because of the very centrality of these professions and their institutions to our society. Ordinary people need to believe in them and resist knowledge that would undermine belief. Over time, they will lose faith, but slowly, almost imperceptibly. *You can hardly see it happening.*

"Personal profit" obviously refers to the professionals' fees and salaries, which normally produce a standard of living that falls somewhere between extremely comfortable and mediocre, but in any case liberates them from the health-uninsured double jobs or double shifts of the working poor, and the wearying physical labor often required of even highly skilled nonprofessionals. (My plumber, who commands a much higher hourly fee than I do, uses muscle and *sweats*. I do not.) Additionally, there are perks of employment that professionals, but few others, can take for granted: office space, personal computers, phone lines, comprehensive insurance, and so forth. But much more significant than any of this—the greatest payment conferred by the client society—is *cultural authority.* This includes social status but is a much thicker, richer, and more problematic concept. It is differentiated by Paul Starr from *social authority,* which "involves the control of action through the giving of commands"; the parent, the coach, the police

officer, and the IRS all have social authority, at least in theory. In contrast, cultural authority, in Starr's words, "entails the construction of reality through definitions of fact and value."[8]

The construction of reality through definitions of fact and value. This is the awe-inspiring nature of the authority that our society vests in the professions it has generated and endorsed, and in those people who have the resources, the talents, the qualities of character, and the general good fortune to enter them. A self-aware professional with even a dollop of humility must, upon reflection, find that authority quite wonderful and terrible, especially in light of the way it is maintained.

Partly, of course, this construction is maintained because the resulting reality seems authentic and serviceable to its constituencies. It persuades. But even more significant is every profession's dogged self-protection—which is justified as being essential to the client's protection but just as clearly serves the interests of the professions themselves. To require degrees, certifications, and licenses is to deny entrance to those who do not have them, regardless of their knowledge base, and, at least potentially, to keep the numbers of professionals from multiplying beyond demand and to keep their market value comparatively high. A system of admission and exclusion is obviously necessary—fundamental to the very idea of expert socialization—but that we have been empowered to do it, to limit and exclude on our own terms, piles very heavy weight onto the burden of responsibility we carry.

Jack and Joe and Jenny and I and most of this book's readers belong to what is a profession writ large. Defined *thick*. The professoriate constitutes a profession *and* a metaprofession, controlling, as it does, not only its own ranks and what students will learn in order to prepare for thoughtful living and useful citizenship, but also what they must know to enter virtually all of the other professions in America and how attainment of that knowledge will be measured. Who gets in—for the benefit of society. Who is kept out—for the protection of society. Who is distinguished as "excellent." Our responsibility is multiplied vastly because what we do affects every one of our ultimate client's professional institutions and activities. Suitably, perhaps, the perks portion of our personal profit is similarly multiplied. The tenured professor, for all practical purposes, has no boss. Her schedule, though busy, is largely under her control: *completely* so for four or five months of the year. How she will teach and in some cases what she will teach are largely up to her, as is the direction her research will take. If she

even does research anymore. And what campus service contributions she will choose. If she even gives service. She can forego stockings and makeup and wear the same granny dress day after day. Her male colleague across the hall can grow a shaggy ponytail, a long Santa Claus beard, and wear cutoffs with lumberjack shirts. No one will say a thing.

No one will say a thing. Because we hold and strenuously guard sovereignty over abstract knowledge and many applications of that knowledge, and complex skills that are essential to society's well-being and growth. Our construction of reality is accepted, and our associated definitions of value are endorsed, to the extent that our definitions of fact are trusted. And, short of the academic equivalent of an Enron or an earthquake, they *are* trusted, more or less. To reject them would be to invite the earth to shudder under our collective feet, the entire social infrastructure based on expert services possibly to collapse. Should we betray that trust less spectacularly, in small ways day by day, grade sheet by grade sheet, trust will erode gradually. *We will hardly be able to see it happening.*

Like it or not, the grades that we assign, which lead to our profession's giving or withholding its imprimatur in the shape of a diploma, constitute our most conspicuous communication as teaching professionals to our clients at each remove: students, students' families, and the public masses who believe us, who have no real choice but to believe us—and to hire and pay our graduates to teach their children, write programs for their computers, design their skyscrapers, interpret their laws, keep their accounts, strategize their wars, report truth in their newspapers, heal their physical and mental ills—in short, to do the most rewarding, perhaps the most interesting, and probably the most critical, work that their society offers. Work that should be performed to the very highest standard, a standard of excellence.

Academics' overuse of the A makes it impossible to distinguish excellence from bare adequacy and tells barely adequate students that their work is excellent or good enough to be *rated* excellent. One Internet apologist for grade inflation claimed that since "everyone" understands the issue (certainly a questionable claim), even the complete disappearance of C, D, and F would pose no problem. The new equation, he argued, is: A equals A; A– equals the former B; B plus is the former C; B is the former D; and B–, of course, is the old F.[9] This sounds silly but seems to be not far off the mark in some institutions and disciplines. It may even be conservative as regards grading in

many graduate programs. It completely misses the point that even if "everyone" did understand the shifted scale, the system's screening function would be disabled. B– is passing, while F is/was not.

What, then, does a baccalaureate degree signify to society today? That the student completed some forms and won admission, that someone paid his tuition, that he attended at least some classes and probably sat for exams? What else? Who knows? Certainly not *necessarily* that he learned the basics of a discipline. And if his cumulative GPA is somewhere around the minimum for graduation, which in many institutions is still the old "average" C, then the degree does not necessarily signify that he learned anything at all. It may well be that we would achieve the same substantive results for our clients, and would save everyone a lot of time, if we *sold* degrees at the C transcript level. Pay your money, get your diploma. To graduate with a B or A average, you would actually have to go to school. I am being facetious in a sense—I know that it will not happen—but I am absolutely sincere when I say that in many programs within many institutions, there would be no substantive difference, no inherent difference in results, between doing that and continuing our current practice where the F and D are used rarely or never, and the C is reserved for deeply defective performance.

The degradation of the degree is well understood by most academicians, and I have not even touched on the wide perception that, in addition to inflating grades, we have in many cases adulterated content and reduced workload. That the baccalaureate does not mean what it once did—that, in fact, it may assure graduate schools and prospective employers of almost nothing—cannot help but become known more widely. And what of our *first*-line client, the student? What does it signify to her? If she still believes that the degree means what it purports to mean—significant learning in the broad liberal arts and foundational mastery of a discipline—then she may be sadly misled about what she has achieved and is qualified to do. If she has come to view us cynically, she will *hope* that it means what it purports to mean, but she will not *know.* Or, she will simply conclude that we are liars.

Inflated grades are lies.

We lie as surely as the doctor would lie if she gave a patient the mild diagnosis he wished to hear rather than the true, harsh one that might, however, lead to his seeking remedy. As surely as a social worker would lie if he overlooked substandard conditions in a home because

the foster parents were basically well intentioned and were *trying really hard.* And as surely as anyone would lie by falsifying to free her own time, to relieve pressure on herself, to endear herself to those unwilling to hear the truth. I have not summoned examples of those last three cases because it is hard to think of a category of professional that fits. Except, perhaps, for politicians—if we consider politics a profession—and educators.

Whatever explanations we give for lying with grades, whatever excuses, the fact remains that we often lie to enhance personal profit in some way. It may be, literally, to protect our incomes: in the hope of burnishing student evaluations or increasing student enrollments in our classes so that we will obtain a tenure-track slot; or, having obtained it, tenure; or, having obtained that, promotion. Sometimes we simply wish to make life more comfortable, with fewer student hassles, more student affection, less time spent on conscientious grading defended in agonizing detail, and better odds on filling sections of narrow or eccentric courses that reflect our research hobby horses and that we yearn to teach. Many of us, unpopular nerds since grade school, cannot bear unpopularity for life. Or, praised as prodigies from kindergarten straight through doctoral studies, we have become as resistant to criticism as our students are. "Faculty want to be *liked*," a colleague reminded me recently, sounding eerily like Willy Loman.

There has been no academic Enron. The indiscriminate high grading at Harvard and Princeton that made headlines in the mainstream press was not news to most people in higher education or to any Harvard graduate: there had been a detailed exposé nine years earlier in the university's alumni magazine.[10] But the news was accompanied in both cases by avowals of determination to reform, and anyway, Harvard and Princeton are, well, Harvard and Princeton: always seen as very different from the ordinary world at the foot of the mountain. But the remarkably homogeneous transcripts of college graduates overall, and their frequent inconsistency with the knowledge and the speaking, writing, and analytical skills that the graduates actually demonstrate, do not go unnoticed.

At the primary and secondary school levels, grade inflation, "social promotion," and other forms of leniency and misleading assessment resulted eventually in the reliance on externally developed examinations and politically mandated standards that teachers so resent and that have forced them to "teach to the test." A disinterested observer may regret this development but does understand what motivated such

interference by outsiders. What, really, does a high school diploma mean now in itself?

*Post*secondary grade inflation has not so visibly eroded educators' professional authority, but the eventual result must be overreliance on standardized tests for graduate school admission and on employer-designed tests for hiring, and ever more emphasis on elitist clues: the prestige of the baccalaureate institution; the quality of the applicant's "connections"; and extracurricular involvements and glamorous enhancements such as volunteer internships and study abroad that may have more to do with the student's resources—for example, that he did not have to spend time at a job in order to pay his tuition—than with his academic ability. We professors should be as appalled and offended by the hegemony of ETS—first SAT, then GRE, LSAT, GMAT, MCAT—as a doctor would be if her diagnoses were routinely supplemented or even overridden by an anonymous third party of unknown qualifications who had never met her patients. And yet we are not. Why?

A search for the answer leads unavoidably to a more painful question: Do we believe in our profession insofar as it involves teaching? Most faculty, after all, do spend most of their working time on teaching, not research. Do we believe in ourselves as professionals? Do we prize our credibility? Can we stand by the "definitions of fact and value" that are being projected through our "construction of reality"? Are we convinced that what we do has critical consequences for our students and our society? And does the quality of these consequences mean more to us than *anything else*? A profession has an ethic of service that teaches that dedication to a client's interest takes precedence over personal profit—which includes personal comfort or ease or happiness—when the two happen to come into conflict.

Can the teaching professoriate honestly be categorized as a profession today?

And do we care?

NOTES

1. The College of New Jersey links grade summaries to its home page through the Office of Institutional Research's portion of the site. The summaries begin with Spring Semester 2001.

2. For the sake of familiarity and simplicity, I am using this term. Elsewhere in this volume, and in two periodical articles, Richard Kamber and I have

argued that it is dangerously misleading, and we have proposed grade *con*flation as a more accurate alternative. See Kamber, "Understanding Grade Inflation," in this book; Richard Kamber and Mary Biggs, "Grade Conflation: A Question of Credibility," *Chronicle of Higher Education* (April 2001); Richard Kamber and Mary Biggs, "Grade Inflation: Metaphor and Reality," *Journal of Education* (Spring 2004).

3. Numerous studies have documented grade inflation at individual colleges and universities or at small groups of institutions—the largest part of this inflation apparently occurring before the latter 1970s, when grades began to slope upward more modestly or flatten, but not decline. Other chapters in this book cite many of these studies. Rather than repeat these citations, I will simply note four recent sources that summarize many of the data: J. E. Stone, "Inflated Grades, Inflated Enrollment, and Inflated Budgets: An Analysis and Call for Review at the State Level," *Educational Policy Analysis Archives* 3:11 (1995); G. D. Kuh. and S. Hu, "Unraveling the Complexity of the Increase in College Grades from the Mid-1980s to the Mid-1990s," *Educational Evaluation and Policy Analysis* 21 (Fall 1999): 297–320; H. Rosovsky and M. Hartley, "Evaluation and the Academy: Are We Doing the Right Thing? Grade Inflation and Letters of Recommendation," Occasional Paper (Cambridge, MA: American Academy of Arts and Sciences, 2002); V. E. Johnson, *Grade Inflation: A Crisis in College Education* (New York: Springer-Verlag, 2003).

4. See, for example, K. W. Arenson, "Is It Grade Inflation, or Are Students Just Smarter?," *New York Times: Week in Review* 9 (April 18, 2004).

5. See, for example, J. R. Young, "Homework? What Homework?," *Chronicle of Higher Education* (December 6, 2002): A35–A37.

6. Johnson has prefaced his own research, which was conducted at Duke University, by summarizing all the other substantive studies of the relationship between faculty grading and student evaluations.

7. B. J. Bledstein, *The Culture of Professionalism: The Middle Class and the Development of Higher Education in America* (New York: Norton, 1976).

8. P. Starr, *The Social Transformation of American Medicine* (New York: Basic, 1982).

9. Personal communication.

10. C. Lambert, "Desperately Seeking Summa," *Harvard Magazine* 95 (May–June 1993): 36–40.

CHAPTER SEVEN

Fissures in the Foundation: Why Grade Conflation Could Happen

MARY BIGGS

INTRODUCTION

The ground that yielded so easily to upward grading pressures a few decades ago may actually have begun to soften much earlier—with the advancement of empirical science and statistical techniques in the late nineteenth and early twentieth centuries, when educators became intellectually and emotionally invested in the goal of provably accurate performance testing and grading. The inevitable disappointment of that unrealistic expectation would have coincided with the rise of psychological science and the "progressive education" movement. The emphasis slowly shifted from grades' accuracy to their "fairness" and effects on students' feelings. Their purpose became much less clear, their application much less consistent. I argue that in this shift, and in the confusion of beliefs that both underlay and followed from it, may be found the necessary precondition for grade conflation. When a variety of pressures converged in the mid-twentieth century, they trampled and reshaped this softened ground into paths that led to places where few educators had really wanted to go or had seen looming until too late.

Permanent solution of the problem will require not only addressing discrete immediate factors but engaging in profession-wide discussion of the essential purpose of grades, with a commitment to reaching agreement and radically restructuring grading practices.

Many writers have hypothesized, or even claimed certain knowledge of, the "cause" of rising grades, but every such analysis has been simplistic or superficial at best, biased at worst, and lacking historical context. My interest is not in pinpointing a specific cause or even in

describing a constellation of causes. Rather, I seek to understand how it was that immediate factors, whatever they were, could have had so deep and devastating an effect. Why were grading systems and practices so soft, so vulnerable to distortion? A historical review of the literature of education suggests some interesting answers. First, however, it may be useful to summarize briefly the immediate causes that have been suggested by other authors.

IMMEDIATE CAUSES HYPOTHESIZED IN THE LITERATURE: A SUMMARY

Irrationally rising grades were becoming apparent by the late 1960s, though there is evidence that they had been rising for a decade or more. Not surprisingly, many commentators sought the origin of this change in the circumstances and ideas of the Vietnam War era. Most obviously, faculty resisted failing a student who might thereby end up in Vietnam, his student draft deferment cancelled. But the *ideas* of the period are often seen as even more significant. These include, for example, "New Left" and human rights ideologies that denigrated authority, especially white male professional-class authority. Some faculty became uncomfortable with their traditional role of "judge" and the power it conferred, sometimes leading them to question the basic value and justice of all grading systems.[1]

Strengthening the impact of these ideas were the new groups of students who, by the seventies, were beginning to appear where they had rarely or never been seen before because of their class, race, sex, ethnicity, or age. Some writers argue that faculty eased standards for them (for example, by taking "effort," rather than just achievement, into account), which led inevitably to the easing of standards across the board and the quiet de facto creation of new norms.[2] However, that interpretation is historically implausible, chronologically impossible.[3]

Also noted is the innovative spirit of the time, which may have affected GPAs without conscious alteration in the grading practices of individual instructors. Writers point to "experience-oriented" and independent study courses; college credit for "voluntarism"; the increased popularity of courses and majors long associated with lenient grading; the expansion of electives relative to requirements; the vogue for "criterion-referenced" grading; and thinned course content with reduced student

workloads.[4] Sometimes, that is, higher grades may have been justified, but only because of lowered expectations.

Other writers have also described a changed dynamic between faculty and students. They believe that a shift in the distribution of power between students and faculty arose from new, or more intensely felt, political and social ideologies, and then advanced by the force of its own weight and the constituencies it had created. The evidence cited includes the introduction of pass/fail grading options and lenient course withdrawal/ course retake policies (producing what David Riesman called "the sanitized transcript" [Riesman 1980, 76]) and an increased faculty willingness to allow exams to be retaken and papers to be rewritten, and to assign "extra credit" work. But SET, or student evaluations of teaching, have been studied and discussed most often.

Though the prospective effects of these were explored as early as the 1930s and 1940s (Blum 1936; Withington 1944), SET took hold and proliferated only in the latter 1960s ("Rate Your Professor . . ." 1966). Soon they became a factor, often heavily weighted, in many institutions' faculty personnel decisions. Empirical data, as well as much informed opinion, have established that easy graders receive more favorable evaluations.[5] Some commentators argue that students sensed they had won the upper hand and began to wield it. Once seen by colleges as children socially and fledglings intellectually, who must be protected and directed by their superiors in age and knowledge, students were transformed into consumers whose demands must be met to hold their "business." And their demands, or at least their demonstrated desires, were for high grades, accompanied by an easily manageable workload and "relevance," narrowly defined, of academic material to the times.

Some writers also note the expansion of higher education in the 1960s and 1970s and an intensified interinstitutional competition for students, suggesting that this may explain why student preferences met so little resistance. But this seems too simple, as grade conflation was extreme at some of the prestigious institutions that have always been able to reject the vast majority of even their *qualified* applicants. Also, there had, of course, been earlier periods of expansion in higher education, and there have always been colleges that struggle to attract students.

Some writers hypothesize a connection between lenient grading and the heightened emphasis on research at all types of institution. This seems plausible: it is much less time- and energy-consuming to give a high grade, which no student will challenge and which need not be justified or even accompanied by a comment, than to give a low one. Still other explanations have assumed an increased incidence and toleration of student cheating, given the ubiquity of term paper vendors and, more recently, the Internet; or they have focused on pressure from administrators and accrediting groups that prize retention over rigor.

Finally, taking a long view, and perhaps also seeking to evade blame, some writers from higher education have discerned a sort of "trickle-up" effect when students are conditioned in primary and secondary schools to expect high grades for little work and are misled about their abilities. In fact, educational theory and practice among the levels—primary, secondary, higher—have never been separable, have always been markedly interactive and mutually influential.

But these writers fail to consider the extent to which lenient grading in primary and secondary schools is *set in motion* at colleges and universities. Decades of data have established that students of education tend to be among the weakest in any college or university, as measured by criteria such as SAT scores and high school rank, while education faculty award the highest grades of all—outdone only by music faculty.[6] Thus socialized, their graduates are likely to repeat the pattern in their own teaching.

SELF-PERPETUATION AND INERTIA

Whatever the initial causes of grade conflation, it seems clear that a major cause of its continuance is the practice itself. Once underway, it establishes its own communities of interest, becomes its own rationale, and is extremely difficult to reverse despite educators' near-consensus about its negative effects. Students, naturally enough, have always wanted high grades, and parents take them as evidence that their offspring are excelling, their schools and colleges are succeeding, and their taxes and tuition dollars are not being squandered. Ignoring pangs of conscience, many faculty find that the practice saves them time, spares them conflict, and may help protect their jobs and advancement. Administrators and department chairs are glad—sometimes

desperate—to attract and hold students. Soon the only meaningful grade is a low one, and low, or even middling, grades are thought to place the students who get them at a disadvantage when competing for jobs and graduate admissions, and to place the colleges that give them at a disadvantage when competing for students. Grade conflation, no matter how damaging and demoralizing, seems necessary to continue once begun.

FISSURES IN THE FOUNDATION

But all of this conjecture and speculation begs a deeper question: How could grade conflation take hold so easily and dominate so remarkably? I have summarized the factors most commonly believed to be significant, and all of these may have had some immediate effect; I dismiss none and do not argue one above another. However, it is clear historically that most of these conditions had been present at various times in the past without releasing grades to bob upward ceiling-high. None of them, taken singly or together, adequately explains the dramatic wrench in how faculty applied grading policies that, on paper, usually did not change.

What had eaten away at the ethical foundation of traditional grading to make it so soft, so susceptible to challenge, by the late-middle twentieth century? I commence the search for an answer to this question by asking: When and why did grading begin? What has been the relationship of graded (student) to grader (faculty)? What, over time, have we imagined the strengths of grading to be, and what attitudes toward grading were being expressed in the decades immediately preceding the 1960s? Looking all the way back to America's first college, colonial Harvard, we find the earliest recorded examination: a 1646 public recitation in Latin.

FACULTY, STUDENTS, AND GRADES IN AMERICA BEFORE THE TWENTIETH CENTURY

At the beginning, in the seventeenth and eighteenth centuries, "grades" were, quite simply, a way to let students know how they were doing and to determine whether their performances merited promotion, graduation, or, perhaps, honors. Most nonacademics, and even a few within the academy, probably assume that today's grades serve the

same purposes—although, as we have seen, the link between grading and communication of the quality of student performance has been greatly weakened, perhaps ruptured altogether.

Initially, as at Harvard in 1646, exams were oral and aimed to assess nothing more than the amount of information the students had packed into their minds (Smallwood 1935, 8, 104–105). Later, using written tests, student writing and reasoning skills were assessed as well. Eventually, exam by recitation would dwindle and disappear, lingering mainly in the dreaded doctoral "orals." According to Mary Lovett Smallwood's history of grading in five early colleges (Harvard, William and Mary, Yale, Michigan, and Mount Holyoke), which has been elaborated upon but not superseded, the first known grading *system* comparable to ours was instituted at Yale in 1785 and ranked students Optimi, second Optimi, Inferiores, or Pejores. Essentially, it was a 4-point system using words rather than single letters or numbers. Approximately 180 years would pass before broad and roughly symmetrical grade distributions across 4- or 5-point scales would skew upward to become unreflective of what common sense tells us about human performance.

In the early 1800s, numerical scores were appearing: in 1813, for example, a 4-point scale at Yale, including decimals as a rough equivalent of our less sensitive "plus and minus" system, and in 1830, a 20-point scale at Harvard, the final grade to be based on a student's scores in subject examinations. Students in those early colleges, like students in ours, were required to have a certain average, or a certain percentage of grades above a given number, to graduate.

In 1895, Harvard invented a three-category "scale of merit": Fail, Pass, or Pass with Distinction. That this resembled Yale's much earlier four-category scale is mildly interesting; that some of today's "progressive" reformers have unwittingly called for almost identical versions of both scales to help defeat grade conflation is fascinating. Those favoring the Harvard-like three-category scale apparently reason that faculty would tend to grade mostly either-or, pass or fail, reserving the highest honor for the truly outstanding students (e.g., Cole 1993; Walhout 1999).

To be sure, Early America was no educational utopia. The student bodies were tiny and homogeneous, the curriculum narrow and fixed, but the challenges we tend to associate with size and diversity also were present (arguably in more modest form) centuries ago. Hofstadter and

Smith's valuable documentary history of American higher education furnishes many commentaries that, like Smallwood's, match the tone and much of the substance of today's critics (Hofstadter and Smith 1961). Charles Nisbet, a Scot who in 1785 assumed what would be an eighteen-year presidency of new little Dickinson College, complained eight years later that students were spoiled by their parents, returned to school late from vacations, and "are generally very averse to Reading or thinking, & expect to learn every thing in a short Time without Application" (Nisbet 1793, 254–55). This aversion was "encouraged by many Quacks in Education [substitute easy-grading faculty and colleges?], & many Scribblers in the Newspapers & Magazines [substitute the Internet, term paper mills, and Cliff Notes?] whose Nonsense is greedily swallowed by Youth because it flatters their Indolence, & persuades them that Learning may be obtained without Time or Labour" (Nisbet, in Hofstadter and Smith 1961, 255).

Joseph Green Cogswell, an American studying at the great German University of Göttingen in 1817, discovered by contrast that "laziness is the first lesson which one gets in all [of America's] great schools," because Americans, who supremely valued "preparation for making practical men," wanted "but few closet scholars, few learned philologists and few verbal commentators." Admittedly, Cogswell was a tough judge, devoting twelve hours each day to "private study and public instruction" and resolving to increase this to sixteen (Cogswell, in Hofstadter and Smith 1961, 261), but other accounts suggest that he exaggerated little if at all.

Six years later, while planning the future University of Virginia, Thomas Jefferson proposed minimal admission requirements and "uncontrolled choice in the lectures [students] shall wish to attend," not for intellectual or pedagogical reasons, but because "discipline" was the problem filling him with the "most dread": "The insubordination of our youth is now the greatest obstacle to their education. We may lessen the difficulty perhaps by avoiding too much government, by requiring no useless observances, none which shall merely multiply occasions for dissatisfaction, disobedience, and revolt" (Jefferson, in Hofstadter and Smith 1961, 267). This tells us something, though nothing we did not already know, about Jefferson's philosophy of human freedom and self-determination, but it tells us much more, and this quite shocking, about the existing condition of student-college relations.

Negative statements about students proliferated throughout the nineteenth century. In the 1820s, Professor George Ticknor of Harvard criticized the college's innovative new elective system as superficial and lacking rigor, and he said that the students saw education mostly as an employment credential (Ticknor, in Hofstadter and Smith 1961)—a complaint that has been reiterated in every succeeding decade. For instance, sixty years later, John W. Burgess, a dean at Columbia University, wrote that students generally derided as useless any study "which does not connect itself *directly* to the practice of some profession" (Burgess, in Hofstadter and Smith 1961, 656). Today's often-heard complaints about students' lack of intellectual interests and their "union card mentality" have a lineage stretching back virtually to the beginning of college education in the colonies.

What was to be done with such students? How were they to be treated and taught? The Yale Report of 1828 famously stated that the college must establish "the *discipline* and the *furniture* of the mind; expanding its powers [substitute today's "critical thinking skills"], and storing it with knowledge ["content"]," insisting that the first task was the more important (Day and Kingsley, in Hofstadter and Smith 1961, 278). This report articulated the ideals of *in loco parentis* and the college as a "family," principles soon endorsed by other writers (e.g., Silliman and Adams, both in Hofstadter and Smith 1961), although in 1842, President Wayland of Brown would argue that they were too dominant, and a higher level of student maturity and self-government should be expected and fostered.

The Yale Report also warned of increasingly strident public demands that practical subjects be taught and "the doors . . . be thrown open to all": contrarily, the "first and great improvement which we wish to see made, is an elevation in the standard of attainment for admission," without which Yale would "only expose [itself] to inevitable failure and ridicule" (Day and Kingsley, in Hofstadter and Smith 1961, 285). Within a few decades, the authors' implicit fear was slowly realized as the elite urban universities like Yale began to admit a few Catholics and Jews, a rare "Negro," and in some institutions—among them, Chicago, Michigan, Stanford, and Cornell, though not, of course, Yale or Harvard—women (Veysey 1965, 271–72).

But long before this, the proper relationships of faculty to student, of both to the institution, and of all to the larger society, were being

pondered and debated. The argument of Francis Lieber, a nineteenth-century German political refugee and scholar, against basing professors' compensation on their popularity with students, bears an eerie relevance to our career-influencing usage of student evaluations: "I have seen students fill a lecture room for the mere sake of entertainment, because the Professor interspersed his lecture (by no means the best of the university) with entertaining anecdotes. . . . Have the greater men always been the most popular among the students? By no means" (Lieber, in Hofstadter and Smith 1961, 299).

It seems that there is nothing new under the educational sun *except* grade conflation.

In the mid- to late-1800s, there was sufficient confusion and controversy over the student-faculty dynamic and the legitimacy of grades, given "individual differences" among students (Smallwood 1935, 77), that the University of Michigan banned Phi Beta Kappa for years in resistance to its differentiating of students based on their grades—but (and this is important to remember) differential grades were still awarded, and the "C" remained respectable. However, Michigan—along with Stanford—actually abolished grades for awhile (Veysey 1961, 63); their faculties fretted that lower standards would follow (Smallwood 1935, 77), and in fact, grading was resumed to protect standards.

Education historian Laurence R. Veysey notes that graduates of early nineteenth-century colleges unanimously judged the accomplishment easy, and Frederick Rudolph agrees that "nowhere were students being dismissed for academic reasons and nowhere were really challenging intellectual demands being placed upon them" (Rudolph 1962, 288). Programs of study matured, faculty improved, and standards stiffened as the century aged, especially with the latter half's university movement (Veysey 1961, 1–18; Tappan, in Hofstadter and Smith 1961). Addressing the Harvard community in 1869, at his inauguration as its president, Charles William Eliot boasted that over one-quarter of the college's matriculants failed to graduate, demonstrating, he implied, its high standards (Eliot 1961, 607). Nevertheless, the establishment of uniform, rigorous standards was hindered by competition among smaller, more vulnerable colleges and even among some universities; by unevenness resulting from new elective systems and vocational curricula; by confusion about collegiate purpose and the roles of

administrators and faculty vis-à-vis students; and by the slow "democratizing" of admissions, among other factors.

In 1965, to support his point that "the usual student of 1900" was "belligerent in his unserious [academic] stance," Veysey reported a 1903 Harvard finding that its average student spent "only fourteen hours a week in study outside the classroom" (Veysey 1961, 272), slightly *more* than American students nationwide in 2000! What was also different was that average Harvard students in 1903 could not count on A or B grades, as they could a century later (Hartocollis 2002). Students' time priorities did not drive faculty grading practices, and systematic grade conflation did not occur.

At the undergraduate level, heavy weighting of grades toward the higher end of the grading scale and virtual abandonment of the scale's center as an average seem to be late-twentieth-century innovations that have carried into the twenty-first century. In earlier periods, the "gentleman's C" was still a C. Those who deplored students' failure to aspire to grades above C were deploring their *mediocrity*, as C was still a middling, rather than "in effect a failing," grade (Riesman 1980, 76). Attempts to eliminate grades were quickly abandoned and did not migrate into subversion of the scale. And Michigan's rejection of Phi Beta Kappa signified a humanitarian refusal, however misguided, to value the A student above the C student. Today, an *avowedly* similar sentiment (though other motives are more powerful) has resulted in rejection of the C itself, and formal honors, though still in existence, hardly recognize distinction when half or more of students receive them (Helmreich 1977; Wolansky and Oranu 1978; Hartocollis 2002).

To summarize, neither student and curricular diversity, relaxed admissions criteria, nor humanitarian progressivist ideals were new in the 1960s. And the negative phenomena often blamed for the grading distortions of recent years—including spoiled, lazy, recalcitrant, inept, and/or "unserious" students focused on grades, credentialing, and work avoidance rather than on learning; faculty doubts about the fairness of grades; institutional competition and pressure from administrators; and low or faltering academic standards—all bedeviled American higher education before 1900, but did not lead to the deep devaluation, the nationwide "grade conflation," that has been seen over the past forty years.

Of course, circumstances changed in the intervening decades: undoubtedly, students became much more diverse, in culturally signi-

ficant ways; perhaps institutional competition grew more intense or administrators grew less committed to academic quality. Or perhaps not. Certainly, higher education was bigger, more complex, more difficult to study and describe, than in 1900, when less than 10 percent of the population even completed high school (Goldman 1985). Still, knowing that none of these "problems" is new, that higher education has always been afflicted by versions of them, inspires one to look more deeply for the causes of grade conflation. It may be rejoined that the coincidence of so many factors *was* unprecedented, though that is questionable. But still, this does not explain why the response was fast and broad grade debasement when other responses were possible.

For example, grades could have been eliminated altogether. (When Michigan and Stanford abandoned grading briefly in the nineteenth century, they apparently did not even contemplate simply transforming almost all grades, Midaslike, into As and Bs.) Grade thresholds for financial aid and Latin honors could have been lowered. Aggressive tutoring and teaching-development programs could have been provided, and faculty research could have been deemphasized in order to demand stronger teaching and advising. (Indeed, faculty could have organized to demand this change; two- or three-section teaching loads, and seven- or eight-month work years, combined with the opportunity to choose one's own research questions—or, after tenure, to explore none at all—certainly could have been attacked in the sixties and seventies as the indefensible privileges of an "elite" class.) Serious peer evaluation of faculty could have been instituted. Variable course scheduling and credit weighting could have been tried. The curriculum could have been fundamentally restructured away from discrete courses—which are, after all, arbitrarily defined. Pedagogy could have been drastically reshaped; despite much talk of innovation, and some action, a mid-afternoon stroll down the corridor of almost any college classroom building discloses room after room with a teacher at the front, talking (sometimes in tandem with a media screen) to seated students, listening (and/or watching).

None of these changes may necessarily have been good (though certainly no worse than grade conflation); here I do not advocate (or oppose) them. The point is that the phenomena of the sixties and seventies could have inspired one or more of these responses universally rather than the one they did. Grade conflation now seems an "obvious"

response only because we are accustomed to it; once upon a time it would have represented an unthinkable moral collapse.

THE TWENTIETH CENTURY BEFORE THE SIXTIES: THE ILLUSION OF SCIENTIFIC CERTAINTY

Reviewing the literature, one realizes that faculty complaints about student values, attitudes, motivation, and performance are ubiquitous in all times and, of course, are consistent with faculty members' responsibility to evaluate. The published commentary on grading is less predictable and more interesting as it struggles to find its footing on the undulating ground of attitudes toward the "sciences" of psychology, intellectual measurement, and statistics.

Evolutionary biologist Stephen Jay Gould summarized the early history, fallacies, and misuses of modern intelligence testing, beginning with Alfred Binet's first test, published in 1905. A trained tester, working directly with the child being tested, would assess performance of basic intellectual tasks arranged by level of difficulty. Later versions of the test keyed the tasks to age level and assigned a score supposedly denoting "mental age," or "Intelligence Quotient." Binet's aim, Gould emphasized, "was to identify in order to help and improve [children], not to label in order to limit" (Gould 1981, 152), and he warned teachers against "the unwarranted pessimism of their invalid hereditarian assumptions" (153). But professionals were already seduced by science, which seemed to offer final answers and elegant solutions to a raft of disparate, messy, previously intractable problems. Binet's testing seemed to promise that one rather easily derived number could place any person specifically, unquestionably, and unchangeably on the grand scale of intellectual potentiality. In 1916, Lewis M. Terman, a Stanford University professor who dreamed of testing all children in order to determine their appropriate places in society, revised and standardized Binet's test into a written exercise that could be completed independently, renamed it the Stanford-Binet, mass-marketed it, and grew rich (Gould 1981, 176–77).

If "intelligence" could be so precisely measured, then why not performance on all intellectual tasks? At last, the tormented issue of grading could be settled and put aside! Education writing in the twenties and thirties was strongly influenced by ideas underlying the new

enthusiasm for "IQ testing" and by statistical theory, especially the "normal distribution" or "curve." At the same time, however, educators were drawn to the developing, and related, science of psychology, which raised questions about the effects of grading, especially on young people's self-esteem.

By the twenties, intelligence testing was becoming widespread; there was a new determination to scientize all educational evaluation and confidence that it could be done. Howard J. Banker, writing in the prestigious *Journal of Educational Research,* regretted the problem of teacher inconsistency in grading but asserted that in the long run the "actual distribution of teachers' marks has been repeatedly shown to approach the normal-probability curve" (Banker 1927a, 160). He noted that elementary school students' grade medians declined for the first few years; reached their lowest point in the fifth grade, when the weakest students began dropping out; then began to rise again—which convinced him "that the teachers' marks in these schools have on average clearly reflected the changing mental constitution of the student body" (Banker 1927b, 284). Aside from the smug circularity of his argument, what catches the attention is his belief that school performance grades should and did reflect "mental constitution"—that is, intelligence—which, as people believed then (and many still believe now), had been proven definable, fixed, and scientifically determinable by Binet, Terman,and others.

In the same year, high school principal R. O. Billett wrote that a wise school administrator should mandate "scientific methods of examining and distributing school marks," using two invaluable recent innovations: "the objective examination" and "the normal distribution curve." Teacher preferences for "old-time, essay-type" subjective exams were mere "superstitions," he said, useless artifacts of an antirational dark age that had produced "a ridiculous variation" in grades. As the normal distribution of human ability was now known to be "a fact, not a subject for debate," scientific objective tests would yield scientifically valid scores that would inevitably distribute themselves "normally," and all vagueness and variation would end. To support his point, Billett cited an analysis of 96,853 marks issued at four major universities; the modal grade was C, accounting for 38 percent of the total, with approximately 33 percent and 27 percent grouping on either side. "Literally hundreds" of similar analyses,

using fewer cases, had been performed at all three levels of education, he claimed (Billett 1927, 53–54). Grading uncertainty was a relic of the unenlightened past.

Three years later, M. J. Nelson, of Iowa State Teachers College, noted that there was "considerable agitation" about the "unreliability" of grades and observed that "for administrative purposes at least, it would be convenient to have a rather high degree of uniformity" in grading *systems* if not in grades themselves (Nelson 1930, 67). By surveying eighty-nine four-year colleges and universities, Nelson learned that: "objective tests" were increasingly being used, though only six administrations required them; C was considered the "average" grade; and at forty-nine institutions (or 55%) faculty were given a "recommendation . . . as to the distribution of grades expected"—in every case, based on the statistical "normal curve." (Most often, 5% As, 20% Bs, 50% Cs, 20% Ds, and 5% Fs were "recommended," but there was much variation.) Although faculty sometimes skewed the curve toward the upper end of the range, awarding more As and Bs than Ds and Fs and producing an asymmetric left-hand bulge, C seems always to have remained the mode, A the special distinction, and F a live possibility.

In 1932, Carl C. W. Nicol declared that "a given grade should mean the same thing to all instructors and to all departments in-so-far as this is possible" (Nicol 1932, 21). His analysis of Oberlin College grades given in the periods 1912–1919 and 1924–1927 had disclosed a slight overall elevation in the latter period and, more disturbingly, in both periods, an almost 50–50 faculty split between those whose median grades were B and C. Oberlin eliminated the inconsistency in the period 1930–1931 by introducing a ranking system (though conventional grading was later reinstituted). That this inconsistency was considered a startling revelation, and so problematic that it must be and was attacked and solved (however impermanently), illustrates the difference between Nicol's age, with its aspiration to scientifically precise grading, and ours, resigned not only to grading inconsistency but to grades that have conflated into near meaninglessness.

In a 1934 issue of the *Peabody Journal of Education*, a Kentucky school superintendent published what may have been the oddest of the period's educational genuflections to intelligence testing. He correlated the IQ scores of students in grades four through twelve with teachers' ratings of the youngsters' behavior. Three different tests had

been administered to each student, sometimes yielding strikingly different scores, but this did not bother the author, who simply calculated the median of the three and treated it as the child's proven "intelligence." Arraying long strings of data on a table, he demonstrated "a very definite relation" between children's low scores and teachers' low opinions of children: ". . . a teacher may expect most of the disturbing forms of behavior to come from the lower mental levels of her pupils" (Jaggers 1934, 258).

One year later, Grace E. Carter, principal of the San Francisco State College [Teacher] Training School, and Walter C. Eells, a Stanford University Professor of Education, reported that there was regrettably "no general agreement between institutions, or among faculty members in the same institution, as to the percentage of students which under normal conditions should receive each grade," but stated that "the use of scientific measurement is rapidly growing" and clearly thought it should be based on the normal curve. They noted that some colleges (presumably including Oberlin) were experimenting with systems like the one Nelson had described, in which students were ranked by faculty and the ranks were converted to grades by an administrator applying the normal curve. (If, for example, a 5-20-50-20-5 curve were used for a class of 20 students, then the first-ranked student would receive an A, ranks 2–5 a B, ranks 6–15 a C, ranks 16–19 a D, and the last-ranked student would fail.)

There were, however, "conflicting opinions" as to what grades should signify. The scientific camp favored knowledge as the only criterion and sought to arrive at "an absolute standard of measurement in each subject which may be used by all teachers in giving grades. . . . Such instruments as standardized achievement tests, comprehensive examinations, and new type [i.e., objective] tests" would be used. But an opposing group of educators, concerned above all with stimulating student motivation and emotional well-being, urged recognition of "individual differences" and grading that reflected factors other than achievement: especially "attitude, industry, improvement, and regular attendance" (Carter and Eells 1935, 310–11). This was because a second significant strain of educational thought was emerging, which found its inspiration not in the sciences of testing and statistics but in the developing science of psychology, and here the emphasis was not on standardization but on the unique individual.

PSYCHOLOGY AND PROGRESSIVISM

The twenties and thirties were the decades not only of the normal distribution curve and objective test but of the progressive education movement that began at Antioch College in 1921 under President Arthur E. Morgan, continued to find enthusiastic adherents through the thirties and forties, and has not yet disappeared altogether. The progressives judged traditional education a failure, inflexibly teaching disaffected students, with poor language skills and no sense of social responsibility, material that they found irrelevant. The elementary and secondary schools were without purpose, college education was rigid and ineffective, and the two levels did not communicate. Progressive education, in contrast, centered everything in the student. Its dominant values were "initiative, self-expression, creative work, independence, and self-dependence." At Antioch, for example, which combined a few on-campus classes with off-campus work, each student developed, managed, and followed a personal plan of study. Under Hiram College's "intensive course system," only one subject was studied at a time, sparing the students fragmentation of time and attention. Bennington College, which opened in 1932, blended curriculum with extracurriculum and study with living, and, when selecting its faculty, privileged life experience over scholarly credentials (Rudolph 1962, 473–79). There were several other progressive colleges, but their numbers were very few; their influence, however, was disproportionately strong, and progressivist ideas permeated the academy, finding a fertile ground prepared by disenchantment with traditional education and what was proving to be the false promise of scientific testing.

By 1936, the brilliant, fiercely opinionated Robert M. Hutchins, president of the University of Chicago, declared that colleges and universities in that depression era were being warped in purpose by their relentless competition for students. Although the damage done by penury and greed was emphatically his major point, it was significant that he noted the effects of *both* the scientific and the progressive ideals. On the one hand, passion for "educational measurement" had led to the gathering of meaningless data such as attendance records and scores on regurgitative exams that revealed nothing about "intellectual power" (Hutchins, in Hofstadter and Smith 1961a, 928). On the other hand, "our confused notion of democracy" and "the character-

building theory" had led colleges to admit all applicants, to study what they liked for as long as they liked, and to "claim any degree whose alphabetical arrangement" appealed to them, meanwhile incurring great expense to protect their "physical and moral welfare" (Hutchins, in Hofstadter and Smith 1961b, 929, 927).

Very soon, the optimism once generated by measurement and statistical theory *and* the almost moral passion for education centered around creativity, self-direction, and the hegemony of the individual were both in abeyance. Neither of them alone nor both of them in bizarrely yoked tandem seemed capable of producing motivation in students, success in learning, and accuracy and fairness in assessment.

In essays published just four years apart, the University of Chicago's Mortimer J. Adler promulgated logically inconsistent views. In 1941, he deplored the "frothy and vapid" education being offered in America at all levels below graduate school because teachers and parents wanted to protect childhood from work. This idea, he said, had invaded and debased college and adult education as well. Since real learning is serious stuff—"laborious and painful," fraught with the risk of criticism and failure as well as the possibility of pleasure and success—little if any was taking place. This was not a new idea, of course; eleven years earlier, just before the opening of Antioch but years after elements of progressive education had begun to be introduced into the schools, Abraham Flexner's renowned report on American higher education had called students "coddled": "Every jerk and shock must be eliminated. . . . How is it possible to educate persons . . . who are protected against risk as they should be protected against plague, and who . . . have no conception of the effort required to develop intellectual sinew?" (Flexner, in Hofstadter and Smith 1961, 909).

But by 1945, Adler was celebrating the "democracy" and open access that Robert Hutchins had condemned, and seeming to oppose both choice and prescription ("no departmental divisions, no electives, no separate course in which grades are given for 'covering' a specified amount of 'ground,' no . . . true-false examinations . . . lectures kept to a minimum and . . . of such generality that they can be given to the whole student body without distinction of year"). Basically, he was arguing *against* all conventional demands on the student and *for* the overwhelming significance of the "discipline" of the mind, to use the Yale Report's term, at the almost total expense of the "furniture":

> No student should be dropped from college because he fails to measure up to an arbitrary standard determined by a percentage of mastery of a subject matter or skill; he should be kept in college as long as he manifests any development of his own capacities, and lack of such evidence should be interpreted as a failure on the part of the college, not the student. (Adler, in Van Doren 1988b, 112–13)

Although his 1945 essay did not directly contradict his 1941 essay in theory, Adler's tone and fundamental perception of the student were different enough to confuse any reader seeking guidance from his *oeuvre*. But then, it was a confusing time.

GRADING CONTROVERSIES AND CONFUSION: THE SHIFTING GROUND

By the forties, letter grades were so widely used as to be effectively universal, yet controversies about grading were growing ever more strident and often touched raw nerves. Although some writers continued to chase the dream of scientific objectivity and precision in grading (e.g., Hull 1939; French 1951), and adherence to the normal curve would not be totally abandoned for some years, many people no longer believed that thorny questions about grading could be answered by *any* science.

In 1947, President Truman's Commission on Higher Education for Democracy noted the influx of veterans into higher education and criticized colleges for recognizing only that type of intelligence characterized by "verbal aptitudes and a capacity for grasping abstractions." Colleges and universities were too intellectual, the commission wrote, and should also foster and reward such "aptitudes" as "social sensitivity and versatility, artistic ability, motor skill and dexterity, and mechanical aptitude and ingenuity" (Zook et al., in Hofstadter and Smith 1961a, 979). (Predictably, Robert Hutchins shot back a rapid and belittling response [Hutchins, in Hofstadter and Smith 1961a].)

One of the quirkiest essays from this period was written for a general audience by Lyle Owen (who claimed to have taught in both private and public colleges and universities, a few of which he named) and published by the *American Mercury* under the title "The Low State of Higher Learning." This was, of course, many years before desegregation

and affirmative action and nearly three decades before the term *grade inflation* was coined. Owen's tone was dyspeptic, self-satisfied, and cute, but it still seems significant that he described a wrongheaded application of democracy run amok, with the admission to colleges of "almost anybody," the "flunking" of "almost nobody," and the "coddling" of "that half of the students who bestir themselves excitedly about almost anything except stirring their brains" (Owen 1946, 194–95). The unworthy students he exposed were not the lazy, spoiled boys Nisbet had deplored in the late eighteenth century or the rebellious youngsters who had frightened Jefferson into laxity thirty years after that; not the impatient, anti-intellectual credential seekers who dismayed Burgess in the late 1800s or the initially timid trickle of Catholics, Jews, women, and "Negroes" who began to appear in colleges at the same time. Certainly, they were not the "female students" Louis Goldman would scorn as poorly qualified in 1985 and accuse of spurring lenient grading, or the "black students" Harvard's Harvey Mansfield would denigrate in 2001 as academically inferior culprits torpedoing standards. It behooves us to remember that in every period since at least the eighteenth century, one or more categories of student has been vilified as unworthy, a thorn in the sides of professors, and a threat to academic standards. In *every* period. For Owen, in 1946, it was veterans and (especially) scholarship athletes. His twin disillusioning experiences were, first and on a continuing basis, the low quality of the students, and second, a dean's cancellation of his grades when he failed one-third of a class. Owen identified 10 percent, his usual failure rate, as likely to label a professor an "ogre"—though it was obviously tolerated; and, as evidence of his fairness, noted that his 11 percent rate of As in the problematic class was rather high. He said nothing about mid-range inflation (Owen 1946, 194, 197–98). It seems, then, that the statistical curve was a rough norm in Owen's many institutions; he was attacked for violating that norm, not for doughty defiance of what would later be called "grade inflation."

Just four years after Owen's essay appeared, Chester C. Maxey published a report praising Whitman College's new grading system. Among its virtues were greater ease of administration and clearer-cut criteria for passing, but Maxey also noted that Whitman considered it a more "truthful" system, which better encouraged "superior" performance because the highest grade was awarded less often. Thus at least one late-forties college administration had had enough concern about

the rate of top-letter grading to herald, in print, a lowered rate as an accomplishment.

The most conspicuous change of the time—even at Whitman College—seems to have been some reduction in the far right-hand side of the curve, or the low side not departure from the curve altogether. But there were hints by the early fifties that this would change. One group of commentators was arguing now that grades *inherently* lacked absolute value—were profoundly, perhaps irremediably, flawed: inaccurate, inconsistent, ambiguous, invalid, poor motivators, not predictive of real-life success . . . the accusations varied (e.g., Spaulding 1954; Constance 1957). Some said that educators should not falsely simulate precision by employing exact percentages or pluses and minuses; others challenged the virtually universal A–F system as too specific or otherwise unsatisfactory; and there were those who would abolish grades altogether (e.g., Walter 1950).

Another group of writers, focusing on students' psyches, wondered if grades were fair; if, fair or not, they were too devastating in their effects; if (self-contradictorily) standards should be personalized rather than standardized (Ward 1954); if their cost in student happiness and self-esteem was too great. That there *was* such a cost seemed to be taken for granted. In a 1954 article that reads much more like one that might have been written four or five decades later, H. C. Brearley, of the George Peabody College for Teachers, warned that educators had become the captives of their psychological concerns about student neuroses and inferiority complexes and were on their way to extinguishing meaningful grades from elementary school through graduate school. There was reluctance at all levels to fail any student, he claimed, and many high school diplomas had become nothing more than "certificates of attendance," forcing college admissions officers to rely on class rankings and standardized test scores: ". . . the evaluative or selective function of the school seems to be steadily declining in importance," he concluded. This, remember, was 1954. Already, the die was cast.

THE "SIXTIES"

Early in 1960, Donald B. Hunter declared that "almost any time teachers get together," they questioned whether grades were "fair" or "vicious," and he warned that they should not succumb to the "great temptation" to use grades as "rewards" (Hunter 1960, 22–23). In a

retrospectively revealing article published the next year, mathematics professor Guy M. Rose recommended, with liberal intent, that faculty—especially those with experience—be given some leeway in grading. But he obviously assumed that grades would always be based on the normal curve; that he did not state and defend this assumption suggests that he thought there was not yet much deviation from it. His idea of leeway was mandating a grade distribution of 31% As and Bs, 38% Cs, and 31% Ds and Fs—but permitting flexibility in how the 31% slopes would be divided. However, the statistically structured system that Rose was reacting to had probably never been quite as rigid and tidy as his response implied—and in any case, it was already in decline, a victim of competing concerns, confusions, and disillusions.

By 1961, according to the editor of the *Journal of Secondary Education*, there were frequent recommendations that high schools abolish conventional grades. Teachers worried that they provided the wrong kind of motivation: that is, to get good grades rather than to learn as much as possible. Indeed, some bright students shunned the more challenging courses in order to protect their grade point averages. Thus grades directly inhibited learning (Bush 1961).

The validity of the normal distribution, formerly accepted as a mathematical truth, was being challenged when applied to small groups (Greene and Hicks 1961) and criticized for the class-to-class inconsistencies that it engendered. Should an absolute standard replace the curve, which might result in a disproportionate number of high grades for a good class, and many Ds and Fs for a poor one?[7] On the other hand, should a student be measured against any absolute *or* against other students, or only against his own potential? And what should a grade measure? What, indeed, was its purpose? Matters that had seemed quite straightforward in the colonial period and the early decades of the republic had grown much murkier and more complex.

Should all grades be abolished? Or should letters be replaced by some other system (for example, pass/fail or narrative evaluations)? If not, should any student ever receive an F? If so, for what reasons should an F be given? Was it possible to say that a *student* had failed, or was it the *educator* who had failed (to educate)? Were students being overworked? More fundamentally, what were the most precious values, purposes, and goals of higher education itself?[8] It turned out that no one really knew—or, rather, that few people agreed.

Caught up in the mutually reinforcing currents of disenchantment with scientific statistical measurement on the one hand, and anxious preoccupation with softer psychologizing on the other, educators were raising and debating these issues *before* conscription began for service in Vietnam; *before* the Civil Rights Act, affirmative action, and the steepest increases in percentages of women students; *before* the New Left and the "counterculture" influenced academic ideas, curricula, and admissions; *before* the word "feminism" was in common use: in other words, before many of the phenomena that writers have seen as causing educators to question grades and ultimately conflate them into innocuousness.

INTO THE TWENTY-FIRST CENTURY

Today, the end of the debates about general questions regarding higher education seems ever farther away. But by the late seventies, stated concerns about grading fairness and the effects of grades on self-esteem (which, however, is still a distorting issue—as witness Sykes 1995 and Edwards 2000) had been largely replaced by exposés and discussions, almost always negative, of what had first been labeled "grade inflation" only a few years earlier. Uncertainties, disputes, and disappointments that went back decades—exaggerated hopes followed by exaggerated disllusionments—had been resolved in one sense by a volcanic resettlement of the grading scale that was astonishing in its scope, force, and lack of formal planning or approval at any level, by any official group. But this change, so far irreversible, bred enormous new problems, so far insoluble.

Ironically, as grades conflated, the "objective" tests and "scientific" methods that had won and then lost educators' confidence gained in prominence and power. Elementary and secondary school curricula and grades became so inconsistent and hard to assess, and unearned "social" promotion so common, that competency testing was widely instituted. The test results effectively superseded teachers' judgments and forced them "to teach to the tests," which many justifiably complained was poor pedagogy, reductive and dull. College and graduate school admissions officers were forced to emphasize SAT, GRE, and other entrance examination scores when faced with homogeneous transcripts.

A milestone (it seems in retrospect) marking the end of one path, however overgrown and puzzling, was the three-day conference on

college grading systems held at Buck Hill Farms, Pennsylvania, in May 1963. Convened by Howard M. Teaf Jr., of Haverford College, it attracted representatives of fifty-two undergraduate institutions, including liberal arts colleges (both stand-alone and within universities) and three service academies. The proceedings were published the following year in the *Journal of Higher Education*, and not one paper mentioned rising grades or academic standards ("A Conference on College Grading Systems" 1964)—which, beginning less than a decade later and continuing to the present, have been central to commentaries on grading. The emphasis in 1963, at least on Buck Hill Farms for those three days, was still on the "science" of testing and assessment: techniques, scales, statistical reliability.

The subject would never again seem so simple.

Early Cassandras had been treated as was the original Cassandra, with similarly disastrous results. We have seen that in 1954, a full fifteen years before "grade inflation" is commonly said to have "begun," Brearley published an article with a title that asked, "Are Grades Becoming Extinct?" He declared that teachers were grading more leniently for fear of turning "a teen-ager into a neurotic or a delinquent," and even college professors were often giving no Fs at all to undergraduates and no grades lower than B to graduate students. Concluding that "the evaluative or selective function of the school seems to be steadily declining in importance," Brearley predicted that college diplomas' credibility with employers would decline. He stirred no noticeable reaction.

Twelve years later, but still years before there was wide consciousness of grade conflation, college president J. W. Cady published a thoughtful article, which reported survey findings of grade high-skewing in Arkansas, North Central, and Southern Baptist institutions. As and Bs combined accounted for 47.5 percent of grades, which Cady found deeply troubling, as both letters supposedly signified *above average* performance. These elevated grades violated the policies stated in the colleges' own catalogs, Cady wrote, and he complained in frustration that no one was paying any attention (Cady 1966). For years, very few would. Hildegard R. Hendrickson, writing a decade later, revealed that GPAs had been rising nationwide since 1960 (Hendrickson 1976).

Dozens of subsequent studies would disclose that while everybody's attention was turned elsewhere, grades (which, as we have seen, had apparently been in slow ascendance for some time), shot up dramatically

between the late sixties and the mid-to-late seventies, then apparently and almost inevitably leveled off in many, though not all, places, but never substantially declined. Now, at last, attention was being paid, but to little apparent effect.

THE ARGUMENT SUMMARIZED— AND NEXT STEPS

Is it too late, then, to restore meaning to grades and re-engage the traditional responsibilities of higher education to certify competence and screen incompetence? No, but it will require leadership at the national level. Many techniques of remediation have been proposed, and some should be tested. In another article, Richard Kamber and I summarize these and make recommendations (Biggs and Kamber 2005).

There will have to be widespread commitment across all types of institution and partnership with secondary and elementary educators. On this issue, there must be multi-level communication and cooperation, even if grading practices differ. What will be most critical in the long run, however, is intensive discussion, within and among colleges and universities, orally, electronically, and in print, of grading purposes, criteria, and techniques. It was lack of cogent discussion, analysis, and policy enforcement that created the chaotic culture in which we as professionals lost our footing. Our world will never be as simple as it was in 1646, when Harvard stood alone, teaching a fixed classical curriculum to a small group of youths. It will not even be as simple as it was in the hectic 1960s. But institutions are interdependent in many ways that affect students, and if we cannot achieve clarity and some level of agreement on these most basic issues, then any changes will be cosmetic and vulnerable when challenged—and they will be challenged.

Understanding the problem's deepest causes—appreciating their strong, aged, entangled intellectual and emotional roots—is necessary to acknowledging how intensive and radical, how widely collaborative, the approaches to solution must be.

Here I have argued that "the" cause is not Vietnam, or minority (or veterans' or women's or athletes' or nineteenth-century farm boys') admissions, or materialist values, or faculty who would rather do research than teach, or students who would rather be entertained than educated, or even student evaluations of teaching. It is none of them alone and,

in truth, it is none of them significantly. These immediate factors, plausibly causative, may have acted like seeds that fell into fissures in the grading system. Though the seeds may have taken root and wrought damage, they could not have done so—at least, not broadly and radically—if the fissures had not existed.

The preexisting and fundamental cause, it seems to me, was educators' loss of confidence in their ability to evaluate: the loss of confidence that resulted when their overinvestment in an ideal of scientific certainty betrayed them. This led to a loss of faith in their authority, in their *right* to evaluate. Slowly, the foundation of the grading system eroded, and by the late 1970s, it was revealed to be balanced precariously on sand.

What this means for us today is that the true roots of grade conflation are too deep, and the motives of lenient graders too personal, their habits too entrenched, to be reversed by superficial reforms. Serious dialogue must commence, under professional leadership at the national level, regarding the purposes, the practices, and the ethics of grading. The alternative is continued chaos and continued loss of credibility for higher education.

NOTES

1. See, for example, Bloom (1966); Keniston (1966); Curwin (1976); Tiberius and Flak (1999). Among the commentators who have considered or proposed an end to all grading are Marshall (1968); Kavanaugh (1970); Eble (1972); Dreyfuss (1993); Edwards (2000).

2. Among these writers are Longstreth and Jones (1976); McCuen et al. (1978); O'Connor (1979); Kolevzon (1981); VanAllen (1990); Anderson (1992); Mansfield (2001).

3. For a persuasive refutation of charges made against African Americans, see Cross (1993).

4. Among the writers referred to are Carney, Isakson, and Ellsworth (1978); Prather, Smith, and Kodras (1979); Riesman (1980); Oldenquist (1983); Rosovsky and Hartley (2002).

5. For evidence from studies of SET, see, for example, Hoyt and Reed (1976); Kolevzon (1981); Franklin, Theall, and Ludlow (1991); Placier (1995); Sacks (1996); Greenwald and Gillmore (1997); Williams and Ceci (1997); Archibold (1998); Wilson (1998); Landrum (1999); Krautmann and Sander (1999); McSpirit, Kopacz, Jones, and Chapman (2000); Rosovsky and Hartley (2002); see especially Johnson's (2003) book-length study based on a thorough review and summary of

the literature and a sophisticated empirical study conducted at Duke University. For additional informed opinion, see McKenzie (1975); Longstreth and Jones (1976); Renner (1981); Sacks (1996); Basinger (1997); Trout (1997).

6. For research evidence, see Wolffe and Oxtoby (1952); Kurtz and Bertin (1958); Weiss and Rasmussen (1960); Carney, Isakson, and Ellsworth (1978); Kapel (1980); Placier (1995); Brucklacher (1998). At my college, grades are sky-high everywhere, but in education, they reach the stratosphere; see my chapter in this book, "Grade 'Inflation' and the Professionalism of the Professoriate."

7. See, for example, Johnson (1961) and Aiken (1963), who still assumed that the average grade would be C.

8. Examples of authors who confronted one or more of these questions are Constance (1957); Weiss and Rasmussen (1960); Bush (1961); Johnson (1961); Adler (1988b); Aiken (1963); Bloom (1966); Keniston (1966); Meyerson (1966); Marshall (1968); Kavanaugh (1970); Eble (1972); Curwin (1976).

REFERENCES

Adams, J. 1961. "Jasper Adams on the Relation between Trustees and Faculty, 1837," in *American Higher Education: A Documentary History*, ed. R. Hofstadter and W. Smith, 311–28. Chicago: University of Chicago Press.

Adler, M. J. 1988a. "Invitation to the Pain of Learning," in *Reforming Education: The Opening of the American Mind*, ed. G. Van Doren, 232–36. New York: Macmillan.

Adler, M. J. 1988b. "Liberal Education: Theory and Practice," in *Reforming Education: The Opening of the American Mind*, ed. G. Van Doren, 109–13. New York: Macmillan.

Aiken, L. R., Jr. 1963. "The Grading Behavior of a College Faculty." *Educational and Psychological Measurement* 23: 319–22.

Anderson, M. 1992. *Impostors in the Temple*. New York: Simon & Schuster.

Archibold, R. C. 1998. "Payback Time: Give Me an "A" or Else." *New York Times*, May 24, sect. 4, 5.

Banker, H. J. 1927a. "The Significance of Teachers' Marks." *Journal of Educational Research* 16 (October): 159-71.

Banker, H. J. 1927b. "The Significance of Teachers' Marks." *Journal of Educational Research* 16 (November): 271–84.

Basinger, D. 1997. "Fighting Grade Inflation: A Misguided Effort?" *College Teaching* 45 (Summer): 81–91.

Biggs, M., and R. Kamber. 2005. "How to End Grade Conflation." Unpublished manuscript.

Billett, R. O. 1927. "The Scientific Supervision of Teachers' Marks." *American School Board Journal* 74 (June): 53–54, 149.

Bloom, G. 1966. "There Should Be No F's." *Business Education World* 46 (March): 13–14.

Blum, M. L. 1936. "An Investigation of the Relation Existing between Students' Grades and Their Ratings of the Instructor's Ability to Teach." *Journal of Educational Psychology* 27: 217–21.

Brearley, H. C. 1954. "Are Grades Becoming Extinct?" *Peabody Journal of Education* 31: 258–59.

Brucklacher, B. 1998. "Cooperating Teachers' Evaluations of Student Teachers: All 'A's'?" *Journal of Instructional Psychology* 25 (March): 67–72.

Burgess, J. W. 1961. "John W. Burgess' Program for the American University," in *American Higher Education: A Documentary History*, ed. R. Hofstadter and W. Smith, 652–66. Chicago: University of Chicago Press.

Bush, R. N. 1961. "Editorial: The Grading System and Higher Levels of Excellence." *Journal of Secondary Education* 36: 65–67.

Cady, J. W. 1966. "How Important Are College Grades?" *Journal of Higher Education* 37 (November): 441–43.

Carney, P., R. L. Isakson, and R. Ellsworth. 1978. "An Exploration of Grade Inflation and Some Related Factors in Higher Education." *College and University* 53 (Winter): 217–30.

Carter, G. E., and W. C. Eells. 1935. "Improvement of College Teaching: College Marking Systems." *Junior College Journal* 5: 310–14.

Cogswell, J. G. 1961. "Joseph Green Cogswell on Life and Learning at Göttingen, 1817," in *American Higher Education: A Documentary History*, ed. R. Hofstadter and W. Smith, 260–63. Chicago: University of Chicago Press.

Cole, W. 1993. "By Rewarding Mediocrity We Discourage Excellence." *Chronicle of Higher Education* (January 6): B1–B2.

"A Conference on College Grading Systems [Symposium]." 1964. *Journal of Higher Education* 35 (February): Entire issue.

Constance, C. L. 1957. "Let's Think Again about Grades." *College and University* 32 (Winter): 239–40.

Cross, T. L. 1993. "On Scapegoating Blacks for Grade Inflation." *Journal of Blacks in Higher Education* 1 (Autumn): 47–56.

Curwin, R. L. 1976. "In Conclusion: Dispelling the Grading Myths," in *Degrading the Grading Myths: Primer of Alternatives to Grades and Marks*, ed. S. B. Simon and J. A. Bellance, 138–45. Washington, DC: Association for Supervision and Curriculum Development.

Day, J., and J. L. Kingsley. 1961. "The Yale Report of 1828," in *American Higher Education: A Documentary History*, ed. R. Hofstadter and W. Smith, 275–91. Chicago: University of Chicago Press.

Dreyfuss, S. 1993. "My Fight against Grade Inflation: A Response to William Cole." *College Teaching* 41: 149–52.

Eble, K. E. 1972. *Professors As Teachers.* San Francisco: Jossey-Bass.

Edwards, C. H. 2000. "Grade Inflation: The Effects on Educational Quality and Personal Well-being." *Education* 120: 538–46.

Eliot, C. W. 1961. "Charles William Eliot, Inaugural Address as President of Harvard," *American Higher Education: A Documentary History,* ed. R. Hofstadter and W. Smith, 601–24 Chicago: University of Chicago Press.

Flexner, A. 1961. "Abraham Flexner Criticizes the American University," in *American Higher Education: A Documentary History, R.* Hofstadter and W. Smith, 905–21 Chicago: University of Chicago Press.

Franklin, J. M., M. Theall, and L. Ludlow. 1991. "Grade Inflation and Student Ratings: A Closer Look." ERIC Documents 349: 318.

French, J. W. 1951. "An Analysis of Course Grades." *Educational and Psychological Measurement* 11: 80–94.

Goldman, L. 1985. "Betrayal of the Gatekeepers: Grade Inflation." *Journal of General Education* 37:2: 97–121.

Gould, S. J. 1981. *The Mismeasure of Man.* New York: Norton.

Greene, J. H., and C. R. Hicks. 1961. "Do College Class Grades Follow a Normal Distribution?" *College and University* 36: 296–302.

Greenwald, A. G., and G. M. Gillmore. 1977. "No Pain, No Gain? The Importance of Measuring Course Workload in Student Ratings of Instructors." *Journal of Educational Psychology* 89:4: 743–51.

Hartocollis, A. 2002. "Harvard Faculty Votes to Put the Excellence Back in the A." *New York Times* (May 22), p. A20.

Helmreich, J. E. 1977. "Is There an Honors Inflation?" *College and University* 53 (Fall): 57–64.

Hendrickson, H. R. 1976. "Grade Inflation." *College and University* 52 (Fall): 111–16.

Hofstadter, R., and W. Smith, eds. 1961. *American Higher Education: A Documentary History.* Chicago: University of Chicago Press.

Hoyt, D. P., and J. G. Reed. 1976. "Grade Inflation at Kansas State University: Student and Faculty Perspectives." ERIC Documents 160: 018.

Hull, J. D. 1939. "Can Marks Be Made More Meaningful?" *National Association of Secondary-School Principals Bulletin* 23: 85–90.

Hunter, D. B. 1960. "Can't We Do Something about Grades?" *Kentucky School Board Journal* 38 (April): 22–23, 33.

Hutchins, R. M. 1961a. "Hutchins on the President's Commission, 1948" in *American Higher Education: A Documentary History*, ed. R. Hofstadter and W. Smith, 990–2002 Chicago: University of Chicago Press.

Hutchins, R. M. 1961b. "Robert M. Hutchins Assesses the State of Higher Learning, 1936" in *American Higher Education: A Documentary History*, ed. R. Hofstadter and W. Smith, 924–40 Chicago: University of Chicago Press.

Jaggers, C. H. 1934. "The Relation of Intelligence to Behavior in School Children." *Peabody Journal of Education* 11: 254–59.

Jefferson, T. 1961. "Jefferson on the Educational Regime at Virginia, 1823" in *American Higher Education: A Documentary History*, ed. R. Hofstadter and W. Smith, 266–68 Chicago: University of Chicago Press.

Johnson, M., Jr. 1961. "Solving the Mess in Marks." *New York State Education* 49 (November): 12–13, 30.

Johnson, V. E. 2003. *Grade Inflation: A Crisis in College Education.* New York: Springer-Verlag.

Kamber, R., and M. Biggs. 2002. "Grade Conflation: A Question of Credibility." *Chronicle of Higher Education* (April 12): B14.

Kamber, R., and M. Biggs. 2004. "Grade Inflation: Metaphor and Reality." *Journal of Education* 184: 31–37.

Kapel, D. E. 1980. "A Case History of Differential Grading: Do Teacher Education Majors Really Receive Higher Grades?" *Journal of Teacher Education* 31 (July–August): 43–47.

Kavanaugh, R. 1970. *The Grim Generation.* New York: Trident.

Keniston, K. (1966). "The Faces in the Lecture Room," in *The Contemporary University: U.S.A.,* ed. R. S. Morison, 315–49. Boston: Houghton Mifflin.

Kolevzon, M. S. 1981. "Grade Inflation in Higher Education: A Comparative Study." *Research in Higher Education* 15:3: 195-212.

Krautmann, A. C., and W. Sander. 1999. "Grades and Student Evaluations of Teachers." *Economics of Education Review* 18 (February): 59–63.

Kurtz, A. K., and M. A. Bertin. 1958. "A Reappraisal of Marking Practices." *Educational Research Bulletin* 37 (March): 67–74.

Landrum, R. E. 1999. "Student Expectations of Grade Inflation." *Journal of Research and Development in Education* 32 (Winter): 124–28.

Lieber, F. 1961. "Francis Lieber on the Purposes and Practices of Universities, 1830," in *American Higher Education: A Documentary History*, ed. R. Hofstadter and W. Smith, 297–300. Chicago: University of Chicago Press.

Longstreth, L. E., and D. Jones. 1976. "Some Longitudinal Data on Grading Practices at One University." *Teaching of Psychology* 3: 78–81.

Mansfield, H. 2001. "Grade Inflation: It's Time to Face the Facts." *Chronicle of Higher Education* (April 6): B24.

Marshall, M. S. 1968. *Teaching without Grades.* Corvallis: Oregon State University Press.

Maxey, C. C. 1950. "Whitman College's New Grading System." *College and University* 26 (October): 87–92.

McCuen, J. T., et al. 1978. "Report of the Commission on Academic Standards." ERIC Document 160.

McKenzie, R. B. 1975. "The Economic Effects of Grade Inflation on Instructor Evaluations: A Theoretical Approach." *Journal of Economic Education* 6 (Spring): 99–106.

McSpirit, S., P. Kopacz, K. Jones and A. Chapman. 2000. "Faculty Opinion on Grade Inflation: Contradictions about Its Cause." *College and University* 75 (Winter): 19–25.

Meyerson, M. 1966. "The Ethos of the American College Student: Beyond the Protests," in *The Contemporary University: U.S.A.,* ed. R. S. Morison, 66–91. Boston: Houghton Mifflin.

Nelson, M. J. 1930. "Grading Systems in Eighty-nine Colleges and Universities." *Nation's Schools* 5 (June): 67–70.

Nicol, C. C. W. 1932. "The Ranking System." *Journal of Higher Education* 3 (January): 21–25.

Nisbet, C. 1961. "Charles Nisbet Complains of Lazy Students and Educational Quacks, 1793," in *American Higher Education: A Documentary History*, ed. R. Hofstadter and W. Smith, 254–55. Chicago: University of Chicago Press.

O'Connor, D. A. 1979. "Solution to Grade Inflation." *Educational Record* 60 (Summer): 295–300.

Oldenquist, A. 1983. "The Decline of American Education in the '60s and '70s." *American Education* 19 (May): 12–18.

Owen, L. 1946. "The Low State of Higher Learning." *American Mercury* 63 (August): 194–200.

Placier, M. 1995. "'But I Have to Have an A': Probing the Cultural Meanings and Ethical Dilemmas of Grades in Teacher Education." *Teacher Education Quarterly* 22 (Summer): 45–63.

Prather, J. E., G. Smith, and J. E. Kodras. 1979. "A Longitudinal Study of Grades in 144 Undergraduate Courses." *Research in Higher Education* 10:1: 11–24.

"Rate Your Professor: Another Nationwide Student Activity." 1966. *CTA Journal* (March): 64–66.

Renner, R. R. 1981. "Comparing Professors: How Student Ratings Contribute to the Decline in Quality of Higher Education." *Phi Delta Kappan* 63 (October): 128–30.

Riesman, D. 1980. *On Higher Education: The Academic Enterprise in an Era of Rising Student Consumerism.* San Francisco: Jossey-Bass.

Rose, G. M. 1961. "Analyzing the Distribution of Class Marks." *Journal of Secondary Education* 36: 83–88.

Rosovsky, H., and M. Hartley. 2002. "Evaluation and the Academy: Are We Doing the Right Thing?: Grade Inflation and Letters of Recommendation." Occasional Paper. Cambridge, MA: American Academy of Arts and Sciences.

Rudolph, F. 1962. *The American College and University: A History.* New York: Random House.

Sacks, P. [*pseud.*] 1996. *Generation X Goes to College: An Eye-Opening Account of Teaching in Postmodern America*. Chicago: Open Court Press.

Silliman, B. 1961. "Benjamin Silliman on the Government of Yale, 1830," in *American Higher Education: A Documentary History*, ed. R. Hofstadter and W. Smith, 306–307. Chicago: University of Chicago Press.

Smallwood, M. L. 1935. "An Historical Study of Examinations and Grading Systems in Early American Universities: A Critical Study of the Original Records of Harvard, William and Mary, Yale, Mount Holyoke, and Michigan from Their Founding to 1900." *Harvard Studies in Education,* no. 24. Cambridge: Harvard University Press.

Spaulding, K. C. 1954. "A-B-C-D-F-I." *College and University* 29 (April): 397–401.

Sykes, C. 1995. *Dumbing Down Our Kids.* New York: St. Martin's Press.

Tappan, H. P. 1961. "Henry P. Tappan on the Idea of the True University, 1858," *American Higher Education: A Documentary History*, ed. R. Hofstadter and W. Smith, 515–45. Chicago: University of Chicago Press.

Tiberius, R. G., and E. Flak. 1999. "Incivility in Dyadic Teaching and Learning," in *Promoting Civility: A Teaching Challenge*, ed. S. M. Richardson, San Francisco: Jossey-Bass. 3–12.

Ticknor, G. 1961. "Ticknor and the Harvard Reforms of the 1820's," in *American Higher Education: A Documentary History*, ed. R. Hofstadter and W. Smith, 269–73. Chicago: University of Chicago Press.

Trout, P. A. 1997. "What the Numbers Mean: Providing a Context for Numerical Student Evaluations of Courses." *Change* 29 (September–October): 25–30.

VanAllen, G. H. 1990. "Educational Morality: A Task of Resisting the Economic Corruption of Academic Excellence." *Educational Digest* 317 (January) 232.

Van Doren, G., ed. 1988. *Reforming Education: The Opening of the American Mind.* New York: Macmillan.

Veysey, L. R. 1965. *The Emergence of the American University.* Chicago: University of Chicago Press.

Walhout, D. 1999. "Grading Across a Career." *College Teaching* 45 (Summer): 83–87.

Walter, J. T. 1950. "Grades: Fact or Fraud?" *American Association of University Professors Bulletin* 36 (June): 300–303.

Ward, L. R. 1954. "On Standards." *College English* 15 (March): 348–50.

Wayland, F. 1961. "Francis Wayland's Thoughts on the Present Collegiate System, 1842," in *American Higher Education: A Documentary History*, ed. R. Hofstadter and W. Smith, 334–75. Chicago: University of Chicago Press.

Weiss, R. M. and G. R. Rasmussen. 1960. "Grading Practices in Undergraduate Education Courses." *Journal of Higher Education* 3: 143–49.

Williams, W. M., and S. J. Ceci. 1997. "How'm I Doing?: Problems with Student Ratings of Instructors and Courses." *Change* (September–October):12–23.

Wilson, R. 1998. "New Research Casts Doubt on Value of Student Evaluations of Professors." *Chronicle of Higher Education* 44 (January 16): A12–A14.

Withington, R. 1944. "On Judgment and Grades." *American Association of University Professors Bulletin* 30 (December): 557–68.

Wolansky, W. D., and R. Oranu. 1978. "Clarification of Issues Relative to Student Grading Practices." *College Student Journal* 12 (Fall): 299–304.

Wolffe, D., and T. Oxtoby. 1952. "Distributions of Ability of Students Specializing in Different Fields." *Science* 116 (September): 311–14.

Zook, G. F., et al. 1961. "The President's Commission on Higher Education for Democracy, 1947," in *American Higher Education: A Documentary History*, ed. R. Hofstadter and W. Smith, 970–90. Chicago: University of Chicago Press.

CHAPTER EIGHT

Grading Teachers: Academic Standards and Student Evaluations

LESTER H. HUNT

I sometimes entertain my nonacademic friends by telling them that at the end of each course I teach, before I compute my students' grades, I pause nervously while I wait to be graded by my students. This process can be described less paradoxically, but surely no more truthfully, as follows. In my department, and as far as I know all the departments at my university, each course ends with students anonymously filling out forms in which they evaluate the teacher and the course. The form includes several questions in which the student is asked to rate the teacher in various respects (such as clarity and organization, availability outside the classroom, and so forth) along a numbered scale (in our case, from one to five); they also are asked one very general question in which they are told to rate the instructor's overall effectiveness. They are invited to write comments on these matters. Mean scores (in effect, grades) are calculated for each question and published by the university. In addition, these student evaluations of teaching (often called SET) are used each year by the committee that decides merit pay increases for faculty. When the faculty member is being considered for advancement to tenure or promotion to the rank of full professor, these evaluation forms are supplemented by faculty evaluation of teaching, in which faculty members visit the candidate's classes and report on her or his effectiveness as a teacher. Except for these two once-in-a-lifetime events, the student evaluation forms are the only way in which we judge the quality of teaching. In other words, teaching is for the most part evaluated by students and not by faculty.

What I wish to do here is explain and defend the following thesis: that such evaluation forms, administered and used as I have just

described, are a very bad idea. The only thing that could justify such a policy would be that all the alternatives are, all things considered, even worse. Among the most important considerations that support this thesis are the distorting effects such a policy has on pedagogical method, especially the effect we can reasonably predict it will have on academic standards and grading policies.

One clarifying remark is probably needed before I begin. There are two radically different purposes for which student evaluations can be used. One is their use, by departments and administrators, to aid in personnel decisions, such as pay raises and promotions. The other is their use by instructors who wish to improve their teaching. No one would deny that such evaluations have a role to play in pursuing the latter sort of purpose. I will not be discussing this function of evaluations here, except to point out that evaluation forms that are developed for personnel purposes are in general not intended to help the instructor become a better teacher and would be poorly adapted to serve that end, compared to forms that are written for the express purpose of improving teaching effectiveness.[1]

THREE PROBLEMS

As I see it, the current system, in which we rely almost exclusively on anonymous student forms in evaluating the quality of teaching for personnel purposes, raises three serious problems. I will call them (1) the epistemological problem, (2) the problem of academic standards, and (3) the problem of the distortion of pedagogical styles.

The Epistemological Problem

First, I think it will shed some light on the relevant issues if we pause for a moment to try to imagine a system that clearly would, if it could exist, measure teaching effectiveness. We can easily conceive of such a system if we consider the way the earliest universities were organized. During the Middle Ages, professors offered lecture series on subjects that students needed to know in order to receive their degrees. The lecture series was just that: they were lectures. The professor did not evaluate the students in his own lecture course. Rather, the students were evaluated by taking examinations that were separate from the course of lectures.[2] In such a system, it would be a fairly straightfor-

ward matter to measure the effectiveness of a professor's teaching. One need only look to see how the students in a given professor's course performed on the corresponding examinations. In fact, an *ideal* method would be to examine the students on the relevant subject both before and after the lecture course, then correlate the degree of improvement (if any) with the instructor whose course the student attended.

The reason this would be an ideal evaluation procedure is obvious. Teaching is a purposive activity, the purpose being (in some sense of these words) *to impart knowledge*. It is successful to the extent that it achieves this intended result. It is equally obvious why student evaluation of teaching effectiveness is epistemically problematic. One can only tell whether the result has been achieved if one already possesses it: the only judge of whether a person has come to know X-ology is someone who knows X-ology. Yet whether a given student is in the position of a competent evaluator—whether the student has come to grasp X-ology—is precisely what is at issue when we are evaluating the student's teacher.

I suppose I should emphasize the fact that I am not making the "elitist" point that students are not competent to judge the success of a course as an educational process. The point is rather that some are and some are not, and that, consequently, their *collective* say-so does not constitute a criterion of teaching effectiveness unless we are entitled to assume at the outset that the course was a success . . . in which case, why are we evaluating the course and its instructor? My claim is not an ethico-political one of inferiority but a logical claim of circularity.

To this problem, which I admit (in fact, insist) is fairly obvious, there is an obvious partial solution, which might be expressed as follows. While the ideal method of evaluating teaching effectiveness would be, as I have said, to correlate the efforts of a given teacher with the outputs of those efforts, this would not be possible in this country without fundamental institutional changes. With a few exceptions, such as law schools, professors in the United States evaluate the work of their own students. Of course, it would be implausible to say that we can evaluate the teacher's effectiveness by looking to whether that teacher judges that the students have learned something from the course, so that the fact that Professor Schmidt gives higher grades than Professor Lopez is evidence that Schmidt's students learn more. Still, though we apparently are not currently set up to measure teaching effectiveness by monitoring its outputs, we can find *proxies* for these outputs.

By "proxy" I mean a factor that has some suitable causal relation with the output, which is such that this relation enables us to treat the magnitude of the factor as good evidence of magnitude of the output—so good in fact that we may measure the factor as if it were the output. If you cannot observe a fire, seeing the billowing smoke it produces might be just as good. In evaluating teaching excellence, such proxy factors include such matters as clarity and organization. And this is just the sort of thing that student evaluation forms are about. Anonymous student evaluations do enable us to measure professorial clarity and organization, and this is more or less as good as measuring the degree of knowledge that the students derive from the course.

To this I would reply, first of all, that such evaluations are not for the most part used to measure such alleged proxy factors. The evaluation forms, as I have said, typically contain a question that requests a global, all-things-considered rating of the educational excellence of the instructor or the course. This is the question that typically displaces the others in importance, employed by departmental budget committees or by students who view the university's published evaluation results as in effect the grade that the professor got for that course. This question measures, and is only meant to measure, the professor's student-approval rating. This is precisely the sort of measurement that raises the epistemic problem of circularity that I have described.

Of course, there are other questions in which the students are asked about specific factors that have some causal relation with the educational aim of the course. But some of these factors raise the same circularity problem that the global question raises. Are the "clarity" and "organization" of the course the kind that conduce to knowledge, or do they represent simplifications and falsifications, in which the richness and diversity of a living discipline are forced into an epistemically arbitrary but user-friendly framework? One is competent to answer such questions only if and to the extent that one has achieved the knowledge that the course aims to impart.

On the other hand, some questions are about factors that do not raise this problem. "Did this course increase or decrease your interest in the subject?" seems causally relevant to the educational objectives of the course, and surely all the students know something about the answer to this one. Further, this question reveals a factor that is causally relevant to the educational objectives of the course. However, I would argue that the answers to such questions cannot be treated as evidentiary

proxies for the course objective (i.e., knowledge), because their causal relevance to that objective may be either positive or negative. For instance, I often teach courses that deal with subjects—controversial moral issues, film, literary narrative, the philosophy of Nietzsche—in which students often have a considerable prior interest. This prior interest is often a nonacademic interest. People often take a course on the aesthetics of film because they have seen movies (who hasn't?) and enjoyed some of them (who doesn't?). Some people find that studying film as a rigorous academic discipline actually diminishes the sort of interest they previously had in film. The process of acquiring knowledge includes elements—such as the quest for clarity, the responsibility to justify oneself at every turn, the constant threat of being proved wrong—that tend to diminish certain sorts of enjoyment and interest. Sometimes a course can diminish interest in a subject precisely because it *succeeds* in imparting knowledge. (This raises interesting questions about whether knowledge is an unconditional value, but they are irrelevant here.) The effect on the student's interest in the subject matter is certainly causally relevant to the course objective, but since the relevance can be negative as well as positive, such interest cannot be treated simply as a proxy for achievement of that objective.

In general, the nonglobal questions on a questionnaire, the ones that seek to identify specific factors relevant to educational success, appear to raise one or the other of these two problems. Either they admit of being causally irrelevant to educational success, so that they raise the same circularity problem that the global question raises, or their causal relevance admits of being either positive or negative. In either case, they seem to be poor candidates for the role of *the* means by which we evaluate teaching excellence.

The Problem of Academic Standards

So far I have argued that student evaluations tend to be poor sources of knowledge about teaching effectiveness. Of course, mere ineffectuality is not positive harm. As far as that is concerned, the evaluations might produce random results, so that in the long run, after many trials, they have approximately the same effects on everyone. Of course, this is not the case. They are not random at all. We can expect them to be extremely good indicators of one variable, namely, student approval of the professor and what the professor is doing. To

the extent that they do so, they probably have other effects as well, perhaps both good ones and bad ones. The following is a reason to expect it to have a bad effect, one that should be obvious to anyone who has ever taught a class.

In any department, the faculty will be apt to follow a variety of different grading practices. An assignment that gets a C+ from one professor might well get a B from another. In such a situation, any reason the instructors might have to seek the approval of students will tend to give the instructors who confer the C+ a powerful reason to relax their grading standards, whether they know about the other professors' grading policies or not. The reason is that they will know something about their own interactions with their students. Some of the students who come to such a professor after being graded by a more lenient colleague will be disappointed by their grade and will need an explanation for the fact that the grade they have received is surprisingly low. Theoretically, there are any number of possible explanations, but every teacher who has been in this situation knows that there is a substantial tendency to explain the disappointing difference (quite sincerely) in terms of error, prejudice, or gratuitous meanness on the part of the disappointing grader.[3] Theoretically, it is possible that the low grader might convince the student that her or his estimation of the student's performance is "right," or that, whether it is right or wrong, it is based on a legitimate pedagogical approach. But obviously this is not always going to happen, and some students will think that there is something defective in the judgment or character of the instructor. People in this position will have to either live with a certain heightened level of student disapproval or back off giving the grades that they know are especially likely to cause such problems. There will be a tendency to take the latter course of action, simply because people normally want to be liked and approved of, and they have an aversion to being the object of heartfelt disapproval. If the instructor knows that this disapproval is translated into rating numbers, which then affect the instructor's income, then the tendency will obviously be stronger.

The problem of academic standards is, it seems to me, crashingly obvious. It becomes (if that is possible) even more obvious if we consider it in the abstract. What we have here is a process in which the function of participant *x* is (in part) to continuously evaluate participant *y*'s performance, where virtually the only means by which we evaluate the quality of *x*'s performance is to ask *y* how *x* is doing. Further, we

do this while the process, potentially ego-battering for *y*, is still going on: we do not even wait until *y* has had a chance to gain a sense of perspective about what *x* is doing to put the matter into context. I cannot think of any other situation in our culture in which we have an evaluation process that fits this abstract profile. At least one reason we do not do this is, as I have said, obvious: since *x*'s evaluating *y* is an essential part of this process, this mode of evaluating x interferes with the process itself. It is important that *x*'s evaluation of *y* should track *y*'s *actual performance*. Anything that gives *x* an incentive to evaluate y in ways that track *y*'s *preferences* is a distraction at best and a source of corruption at worst. And that of course is exactly what we do when we set up *y* as the sole evaluator of *x*.

The Problem of the Distortion of Pedagogical Styles

I think I can best introduce this problem by telling a story. Unfortunately, it is a true one. During the sixties[4] there was a professor of philosophy, William B. Macomber, whom I knew both as his student and as his teaching assistant. Of all the teachers I had, he is the one who had, and continues to have, thirty years later, the most powerful influence on the way I think and the way I teach. He was, it is very safe to say, a controversial teacher. He was most notorious for his Introduction to Philosophy course, which had an enrollment of over 700 students. There was only one assigned reading: one of Plato's "erotic dialogues" (one semester it was the *Symposium*, and the next time it was the *Phaedrus*). The exams were all multiple choice and were meant simply to make sure that students did the readings and attended the lecture. The only other assignment was to write a paper. The lectures were brilliant but otherwise hard to describe. They were a mixture of argument, epigram, and anecdote. The anecdotes were mainly about his own life. His basic thesis was that the ideals envisioned by the ancient Greeks, especially Plato, have never been surpassed, and that our own civilization is, in comparison, denatured and decadent. It has been corrupted in every aspect, but especially in its educational system, by the influence of Christianity. He frequently referred to his own homosexuality, relating it to the homosexuality of Plato and using the very different attitudes toward homosexuality in Christianity and the Hellenic world to illustrate the (in his view) deep divide between these two civilizations. In their papers, the students were to defend their views

on one of the issues touched on in the lectures, and it was expected that in many cases they would of course disagree with the professor.

Like the lectures, student reactions to Macomber are difficult to describe. As I have said, he was controversial: by this I mean that students either loved him or hated him. Someone who is universally loathed is not controversial, no more than one who is universally loved. This of course was no accident. In another of his courses he handed out copies of an essay, by classicist William Arrowsmith, called "Turbulent Teachers: The Heart of Education," to justify he own educational practices. In that essay, Arrowsmith argued that the principal aim of a liberal education, especially in the humanities, is to show the student that "a great humanity exists." Since human consciousness does not normally and naturally have much contact with the ways of thinking represented by the great creators of culture, the function of the teacher must be primarily to go against the grain of our ordinary ways of thinking. Inevitably, this means they must upset us and stir us up. Obviously this is what Macomber was doing. It was widely believed by the faculty in our department that his courses inspired more people to become philosophy majors than those of any other instructor. Partly for this reason, and also because of his having recently published a distinguished book, some of us were confident he would get tenure. He did not, and he never worked in the academy again.

I have often thought of him as an early casualty of the anonymous student course evaluations. At the time Macomber was fired, our department had only been using them for a year or two. All the people who were teaching at that time had developed their pedagogical styles in a completely different regime, in which teaching quality was typically either evaluated by faculty or simply ignored. Some of them were still using methods and approaches that could not well survive in the new system. Those who did not change fast enough would have to face some unpleasant consequences, such as, if one is not already protected by tenure, being fired.

Of course, it would be difficult, after all these years, to show that this is what actually happened.[5] However, what is really important for present purposes is to realize that this is just the sort of thing that *would* happen in a regime of numbers-driven student evaluation of teaching. Arrowsmithian pedagogy is not well adapted to survive in the numbers-dominated academy. The new regime rewards people who can identify,

and practice, behavioral strategies that please students. But that is obvious, and it is not the point I wish to make here. The point is that not all strategies of pleasing others are the same, and the new regime systematically discriminates between such strategies. Some of the things that we do that please others are displeasing to no one. They may not please everyone, but they are inoffensive. Others are pleasing to some but displeasing to others. Macomber was a master of the latter sort of strategy. It is entirely the wrong sort of strategy to be using in the numbers-dominated regime. If one student gives me a 5 on the question about my overall effectiveness and another gives me a 1, then they do not merely cancel each other out and disappear from the world. They average to a 2.5, which is substantially below average in my department. If I make one student loathe me, then I have to get *at least* one student to love me, just to approach the semblance of mediocrity.

As far as the numbers are concerned, the results of the latter strategy are indistinguishable from those of teachers that the students rate as poor or mediocre. And here we confront a curious difference between the old regime and the new one. Before the numbers-based system of teacher evaluation, a person was regarded as an extraordinary teacher if (but not only if) he or she had a devoted following. The students who reacted with aversion to the same traits that the others found fascinating were more or less invisible. They were simply people who came to one of the teacher's courses and did not come back. To the extent that the new system holds sway, to the extent that we only look at numbers, what happens is the reverse of this. It is the following that tends to become invisible.

When such a system is linked to pay raises for teachers, it is obvious that it will result in a massive (if subtle, on the micro-level) change in pedagogical behavior. My point is not that this change represents a shift from a superior style of teaching to an inferior style. It is rather that it constitutes an *arbitrary narrowing* of the array of available styles. Defenders of anonymous student course evaluations sometimes point out that they have virtually done away with a certain obnoxious method of teaching, memorably embodied by John Houseman in the film and television series *The Paper Chase*, in which the professor motivates students to study by humiliating the ill-prepared in front of the whole class. This, I think, is substantially true.[6] I would only point out that it does more than that. It harshly discourages the use of *any* pedagogical

technique that can be expected to be abrasive, annoying, or upsetting to anyone. In the current regime, the most rational course is to choose strategies that are inoffensive.

Surely this feature of our system of higher education is a flaw, and a serious one. To deny this, one would have to claim that all educational methods that are displeasing to a significant part of the student population are, by virtue of this fact alone, bad methods and ought to be discouraged. Such a claim denies the familiar fact of human diversity, that different people learn in different ways. To take the extreme case of Kingsfieldian pedagogy, surely there are *some* students who learn better this way, or such methods would never have existed in the first place. Admittedly, the prerevolutionary system, in which students were in effect often compelled to submit to teachers of the Kingsfieldian sort, was deficient, but so is the postrevolutionary one, in which they are in effect *not allowed* to do so, even if that is what they want. But this is, as I say, to take what for my argument is merely the worst case. There are many possible styles of pedagogy, and the current regime of teacher evaluation only makes good sense under the implausible assumption that the inoffensive is by virtue of that fact better than the offensive.

REPLIES TO OBJECTIONS

Such I think are the main considerations that tell against the current system of teaching evaluation. Of course there will be many objections to what I have said. I will conclude by stating and answering some of them.

Student course evaluations are reliable indicators of at least one factor that is always relevant to success in teaching, and it is one that students clearly do know about. This is a response to my argument in (1). In an article offering a limited defense of student evaluation of teaching, philosophy professor Kenton Machina asserts (emphasis in original): "*Student evaluations (if honestly conducted) basically report the extent to which the students have been reached.*"[7] Machina does not tell us what this "being reached" consists in, though he does add to this assertion by saying that the evaluations measure whether students have been "reached *educationally*" (emphasis added). I would like to know whether this state, of being reached educationally, is an emotional state or a cognitive state: does it consist in feeling something or in believing

something? And what feeling, or what belief, does it consist in? Machina also fails to give any evidence for thinking that this state or condition, whatever it might be, is actually measured by evaluation forms.

Nonetheless, the mere fact that he would state the idea so flatly, without justification or explanation, indicates that it must be extremely plausible to some people, and if only for that reason, it deserves to be investigated. Some light on what it means, or might mean, is shed by a recent book, *What the Best College Teachers Do*, by Ken Bain.[8] This book reports the results of a fifteen-year-long project in which Bain studied the almost 100 of the "best" teachers at colleges around the United States. In explaining how he decided which teachers were the best, he says he had two sorts of criteria. He considered such things as how well the teacher's students did on department-wide exams, the numbers of students who went on to achievements of their own, and interviews with former students on how well the teacher's course had prepared them for subsequent activities. Obviously these are examples of the "ideal" sort of evaluation I discussed earlier: evaluation by results. Professor Bain's account is very sketchy on the matter of which of these sorts of tests he applied and to how many of the instructors he applied them.

Concerning the other sort of criterion, however, he is much more clear: to be regarded as one of the "best" teachers, the subjects would have to be rated very highly by students. His defense of taking high student approval as a necessary condition of teaching success is clearly influenced by Machina's article (which Bain cites). I will need to quote this defense virtually in toto to assure the reader that I am not misreporting it:

> We wanted indications from the students that the teacher had "reached them" intellectually and educationally, and left them wanting more. We rejected the standards of a former dean who used to say, "I don't care if the students liked the class or not as long as they performed on the final." We too were concerned with how students performed on the final, but we had to weigh the growing body of evidence that students can "perform" on many types of examinations without changing their understanding or the way they subsequently think, act, or feel. We were equally concerned with how they performed after the final. We were convinced that if students emerged from the class hating the

> experience, they were less likely to continue learning, or even to retain what they had supposedly gained from the class.[9]

This passage contains two dizzying logical leaps. One is the inference from the fact that students like a course and approve of the instructor to the conclusion that they must thereby acquire some of the love of the *subject* of the course, a love that results in further inquiry. Surely there are any number of reasons for enjoying a course and approving its instructor that are perfectly compatible with a disinclination to pursue the subject further. The other inference starts with the plausible claim that if students hate a course they will be less likely to pursue the subject in the future. Here it is more difficult to say what the conclusion is supposed to be, though to be relevant to the issue at hand it would have to be something to the effect that the more students like a course, the more likely they are to pursue the subject further.

The notion that seems to lie behind the idea that evaluations measure whether students have been "reached" educationally seems to be that since real learning is not merely a cognitive process but an affective one as well, then even if the forms only measure students' emotional reaction to a course, they still measure something that is causally related to the indented outcome, which is knowledge. This is undoubtedly true, but as I said before, it makes all the difference what exactly this causal relationship is. Surely the most plausible way to conceive it is to suppose that some degree of enjoyment is a necessary condition of learning. This is clearly the view of Professor Machina, who explains the italicized sentence I have quoted by saying: "Very well-organized lectures, . . . and the like may all be positive factors of some kind, but if they do not result in reaching the students, they will not constitute effective teaching."[10] This is the sort of relationship that might obtain if, to invent an analogous case, human beings were unable to swallow a pill unless it were sweetened by being coated with sugar. In that case the sweetness of a pill would be a necessary condition of its medical effectiveness. But it would not follow that sweetness is a proxy for medical effectiveness, because the causal connection is of the wrong sort. It does not support us in thinking that the sweeter it is, the better it is: that the best pills are the ones that are most highly rated for sweetness. Similarly, there might be some level of positive affect in an educational experience that is *high enough* for the creation of knowledge to occur. It also could be true that in some subjects the highest

levels of student approval are signs that something educationally corrupt is going on, that the pill has been loaded with medically deleterious empty calories.[11]

The empirical evidence does not clearly show that student course evaluations lower academic standards. This of course is a response to problem (2). Machina argues that though there is a statistical correlation between giving high grades and receiving high evaluation numbers, that does not prove that the high grades cause the high evaluations. They may have a common cause: students learning more.

> If the data show a positive correlation between student opinion of a course and grades received in the course, it nevertheless remains reasonable to suppose those students who received good grades learned more, and as a result viewed the course more positively.[12]

Surely it is not reasonable to suppose this at all. Such a supposition flies in the face of a fact that every student knows: that sometimes the reason the students in one class got higher grades than those in another is that their teacher is an easier grader.

On a deeper level, Machina does have a sound methodological point: the fact that two variables are correlated does not prove which one causes the other, or whether there is some more complicated causal relationship between them. Once a correlation has been noticed, the question is: What is the most plausible, best-grounded *explanation* of the correlation? In (2) I explained the correlation by supposing that students can notice discrepancies in grading, and that they have a tendency, to some significant degree (which may be far from universal), to attribute lower grades to teacher error. It is important to notice that my explanation does not deny the existence of the sort of causation that Machina's explanation describes. On the other hand, his hypothesis is that the sort of causation I describe does *not* occur: high grades do not cause any high evaluations. Rather, he is asking us to suppose that all of the correlation between high grades and high evaluations is due to (1) teachers accurately reflecting student learning in their grading practices (no arbitrary discrepancies from one teacher to another) and, in addition, (2) students accurately reflecting how much they learned in their questionnaire-filling practices (and not being influenced by their grades). Which of these two explanations, his or mine, is more plausible?

In fact, recent experimental results published by statistician Valen E. Johnson may make it unnecessary to rely on such a priori reasoning. In Johnson's experiment, 1,900 Duke University students completed a Web-interface questionnaire evaluating their professors and predicting their own grades in their courses. Freshmen in the study filled out the form both before the fall term was over and after. As in countless other studies, students who expected to get an A were more likely to evaluate a professor highly than students who expected to get a B, and they in turn were more likely to evaluate a professor highly than students who expected to get a C. What was novel about this study had to do with the students who evaluated their courses twice. Students who received grades higher than the one they expected tended to adjust their evaluation of the professor *upward*, while those who received lower than expected grades did just the opposite and made a *downward* adjustment.[13] Obviously there is no way to explain this phenomenon in terms of how much the students (really) learned. Grades (or, more exactly, students' perceptions of their grades) *are* a factor, just as common sense suggests they are.[14]

The use of questionnaires to evaluate performance is much more common than I made it sound in (2). Here the objection would be that something analogous to student evaluation of teaching is in fact very common in our culture. Companies use consumer questionnaires to evaluate new products, politicians test their performance by using public opinion polls, and television shows are renewed or canceled based on Nielsen ratings. How is basing university personnel decisions on student ratings any different? My answer is that it is different in at least two ways. First, in none of the previous examples is it part of *x*'s function to evaluate the performance of the *y* who is filling out the questionnaire. Thus none of them involve the unique potential for distraction and corruption that I have pointed out in the case of teaching evaluation. Second, none of the previous examples are cases of evaluating the intrinsic excellence of *x*'s performance but rather its potential for success in the marketplace. The people who use these questionnaires are trying to find out how many consumers will buy their product, how many voters will vote for them in the next election, and how many viewers will see their sponsors' commercials. Evaluation of teaching is (at least officially) intended to measure the quality of the teaching itself, and not how many consumers (i.e., students) it will please.[15]

The pedagogical styles that are encouraged by student course evaluations are ipso facto *superior to those that are discouraged by them: They are* morally *superior.* This objection, a direct response to (3), is one that I have never heard stated, but I suspect that it is at bottom the real reason why the regime of student evaluations continues to exist virtually unchallenged. After all if, as I have said, problems (1) and (2), which raise the question of whether it is rational to measure and evaluate professorial conduct in this way, are obvious, then it is a real question why we continue to do it. It is a paradox that calls for a resolution. Having posed it, though, at least part of the resolution is readily apparent: student evaluations are not meant to measure and evaluate professorial behavior at all, they are meant to *shape* that behavior. The objection is that only my argument in (3) goes to the heart of the matter, but that it gets the matter precisely wrong. The behaviors that are penalized by teaching evaluations are all bad ways to treat people. They are bad because they are morally bad. Generally, the good ways of treating people are things that they will like, and the bad ways are things that they will not like. This is the idea behind the Golden Rule: Do unto others as you, so to speak, would like them to do unto you. The sort of pedagogy that comes out of this ethic is a nurturing pedagogy, one that supports students' self-esteem and kindly encourages their efforts to learn. For the same reason, it also is the sort of pedagogy that is encouraged by the student evaluation system.

It may well be that this kindly ethic of warmth and nurturance will, as Nietzsche claimed, tend to dominate people who live in a democracy.[16] Notwithstanding that, it is subject to some very serious objections. As I have characterized it, it is a species of utilitarianism, inasmuch as it characterizes the ethically good as consisting in giving people what they want. Utilitarianism is per se a perfectly respectable ethical doctrine. However, this is a very peculiar *sort* of utilitarianism, in that it conceives the good as consisting in giving people what they want *now*, at the time that it is given to them. It is, we might say, *short-term* utilitarianism. Every respectable variety of the utilitarian doctrine is based on *long*-term utility. According to long-term utilitarianism, the familiar cliché of parenting—"You'll thank me for doing this in ten (or five or twenty) years"—should (whenever it is true) win every argument. Most utilitarians would probably agree that the ultimate standard for deciding the duty the teacher has toward his or her students is the students' preferences—but they would maintain that the genuine moral standard

is the satisfaction of student preferences in the long term. It certainly is not the satisfaction of the preferences they have when they are eighteen years old, and while the course is still in progress. It is certainly very tempting to treat others in accord with their current preferences, but whenever these preferences conflict with their most likely long-term interests they are, according to the sounder sort of utilitarianism, *only* a temptation: it is our duty to act contrary to them.

Such words as these do leave rather a bad taste in the mouth, I admit—they seem paternalistic. They require us to see students' present selves as incomplete and deficient, and so they seem to speak of them with disrespect. But consider for a moment what the alternative is: to treat people's presently occurring desires as the ultimate standard of value, and to treat offending others as the cardinal sin. This is to deny them the possibility of growth, and to deny the openness to pain and disappointment that their growth requires.

NOTES

1. I am most curious to know for purposes of improving a course such things as which reading assignments the students liked the most, or the least, or which exercises seemed pointless to them. For good reasons (some of which are obvious) personnel-oriented evaluations never ask such questions. Personally, my advice to teachers who want to use evaluation forms to decide how they may improve their educational approach is to design your own forms, with questions designed to provide the sort of information you wish to have, and distribute them in addition to (or even, if permitted, instead of) the administratively dictated forms.

2. Roughly, this system, I am told, is still commonplace in Europe. See Harry Brighouse's discussion in his chapter in this book, "Grade Inflation and Grade Variation: What's All the Fuss About?"

3. A widely accepted theory in social psychology ("self-serving attribution error") holds that people tend to attribute their successes to themselves ("I got an A!") and their failures to others ("She gave me a C!"). See D. T. Miller and M. Ross. "Self-Serving Biases in the Attribution of Causality: Fact or Fiction?" *Psychological Bulletin* 82 (1975): 213–25.

4. As many people have pointed out, when we say "the sixties" we generally are referring to the years 1965 to 1975. In this case, the years involved were 1969 to 1972.

5. For whatever this information might be worth, I recently asked him about the evaluations he got in those courses, and he said that all he could remember was that "they were dreadful," and that they were noticed by the people who had control over his tenure decision.

6. According to a legend, which can still be found on the Internet, when the book on which the film was based first appeared, several Harvard law professors *boasted* that the Kingsfield character was based on them. Today, it is almost impossible to imagine someone bragging about a thing like that—even at Harvard.

7. Kenton Machina, "Evaluating Student Evaluations," *Academe* 73 (May–June 1987): 19–22. The passage quoted is on p. 19.

8. Ken Bain, *What the Best College Teachers Do* (Cambridge, MA: Harvard University Press, 2004).

9. Bain, *What the Best College Teachers Do*, 7.

10. Machina, "Evaluating Student Evaluations," 19. I should point out that Machina's defense of evaluations is very limited, and he in fact argues against the crude, numbers-driven reliance on them that is defended by Bain. He says, for instance, that it would be a "serious error" to use student evaluations to judge the intellectual quality of a professor's lectures.

11. This is probably true of some branches of philosophy. Philosophy often takes the very ideas we live by—God, freedom, immortality—and subjects them to ruthless logical criticism. If it is done right, it is very disturbing, a sort of gambling with one's way of life. Some amount of negative affect is inevitable, unless the professor is falsifying the subject by cushioning students against the shocks—and there are many tempting ways of doing that. No doubt, similar considerations would apply to a number of other humanistic subjects as well, in addition to philosophy.

12. Machina, "Evaluating Student Evaluations," 20.

13. The precise nature of Johnson's findings is complex and difficult to summarize. Those who are interested are invited to consult his book *Grade Inflation: A Crisis in College Education* (New York: Springer-Verlag, 2003), especially 85–118.

14. I should add that everything I say here is consistent with the argument of Clifford Adelman's contribution to this book (Chapter 2), which finds no evidence of consistent grade inflation over the past three decades in American higher education. All of the data he assembles are post-1972, which means that they cover, in my terminology, the "postrevolutionary" period only.

15. Here is one way of appreciating the unique status of anonymous teacher evaluations in our society. They are the only example in our sociopolitical system of power that is routinely exercised in the complete absence of responsibility. Politicians are subjected to checks and balances, judges to judicial review, businesspeople to regulation and the discipline of competitive markets, and witnesses in court are cross-examined, but because the evaluations are anonymous, there are never any consequences to filling them out in an irresponsible way.

16. This at any rate is how I interpret Nietzsche's celebrated discussion of "the last men" in *Zarathustra*. See *The Portable Nietzsche*, trans. and ed. Walter Kaufmann (New York: Penguin Books, 1976), 128–31.

This page intentionally left blank.

CHAPTER NINE

Combating Grade Inflation: Obstacles and Opportunities

RICHARD KAMBER

WHAT GRADE INFLATION IS

To combat grade inflation, one has to understand what grade inflation is. In "Understanding Grade Inflation" (Chapter 3 in this book) I examine various conceptions and misconceptions about grade inflation in order to clarify its nature and consequences. I argue that grade inflation is best understood as a reduction in the capacity of grades to provide true and useful information about student performance as a result of upward shifts in grading patterns. I point out that the harm done by grade inflation is cumulative. Price inflation can be successfully remedied by bringing it down to a low, steady, and predictable rate. Grade inflation must be halted and its ill-effects reversed. I also examine the data on grade inflation from multiple studies and explain that even the lowest numbers show that grade inflation is epidemic in American higher education. I leave for this chapter the practical question of how to get the epidemic of grade inflation under control and steer American educators toward constructive grading practices.

SKEPTICISM ABOUT REFORM

A major obstacle to finding and implementing effective remedies for grade inflation is skepticism about the possibilities for reform. Even educators who recognize the cumulative harm of grade inflation tend to be skeptical about how much can be done to reverse the damage. In their widely read report for the American Academy of Arts and Sciences, Henry Rosovsky and Matthew Hartley urge piecemeal remedies

for combating grade inflation but state that neither "a fundamental systemic overhaul [n]or return to an earlier day" is a "realistic possibilit[y]."[1]

While I do not share this skepticism, I understand from whence it comes. One source is the recognition that local reform, however earnest, is constrained by national practice. This is one of the few respects in which grade inflation resembles price inflation. To some educators, the efforts of individual instructors and institutions to buck the tide of grade inflation may seem as futile as trying to sweep back the sea. A C+ signified average performance in 1960, but today it signifies below-average performance. It is no easy thing to convince students, parents, employers, graduate schools, and so on that a C+ in your class or at your institution means average rather than below average. Another source of skepticism is the realization that grade inflation is self-sustaining and addictive. The more a teacher, a department, or an institution indulges in giving high grades, the more difficult it becomes to return to rigorous grading. Students who have been nurtured with dependable doses of high grades, first in high school and then in college, know they will find withdrawal painful. Faculty who have grown accustomed to dispensing grades that require little or no justification, win smiles from students, and encourage favorable evaluations are apprehensive about policies that will require them to spend more time connecting grades to measurable standards, explaining the grades they give, and dealing with disappointed students. Administrators and trustees who have made student recruitment and retention key measures of institutional success are reluctant to push for policies that may work against these goals. Educators at every level have bought into the myth that as long as there is good teaching and assessment, grading patterns do not matter.

THE NATIONAL SCENE

In all likelihood, the fastest and surest remedy for grade inflation would be a unified campaign at the national level to persuade colleges and universities to adopt a common set of guiding principles for good grading practices. Most of the problems that have accumulated from four decades of grade inflation could probably be corrected by the adoption of just three principles:

1. No more than 20 percent of undergraduate grades at the A level (i.e., A+, A, or A– or their equivalents)

2. No more than 50 percent of undergraduate grades at the A and B levels combined.
3. The grade "F" is used to indicate failure to meet the standards for a course and not just failure to attend classes or complete assignments.

(I elaborate on these principles at the end of this chapter.)

Yet the sad fact is that the most powerful agencies and organizations in American higher education—the U.S. Department of Education, the national associations for colleges and universities, regional and professional accrediting bodies, and state boards and commissions—have done virtually nothing to reform grade inflation. A few small national organizations, notably the American Academy of Arts and Sciences and the National Association of Scholars, have bucked this tide of indifference by sponsoring research and recommendations on grade inflation, but they are the exceptions.

In December 2001, I wrote to the six regional accrediting associations and asked them the following questions: (1) When you are reviewing an institution of high education for accreditation or re-accreditation, what data, if any, on grade distribution do you require? (2) Do you have standards or recommendations for good grading practices? (3) Is the issue of grade inflation a required point of discussion in your accreditation process? Five of the six told me they collected no data, had no standards, and did not require grade inflation as a point of discussion. One of the six declined to answer my questions, explaining that the topic of grade inflation is "politically charged." I posed the same questions to four professional associations that conduct accreditation reviews of undergraduate programs, the American Chemical Society, the Accreditation Board of Engineering and Technology, the American Association of Collegiate Schools of Business, and the National Council for Accreditation of Teacher Education. Again, I was told that they collected no data, had no grading standards, and did not require grade inflation as a point of discussion. It is surprising that professional associations for data-driven disciplines such as business and engineering do not pay attention to grade distributions. It is alarming that attention to grades and grade inflation issues is not part of the accreditation process for teacher education. One would expect teachers of teachers to be especially concerned about such matters, since grade inflation is a serious problem in secondary education,[2] and schools of education tend to give higher grades than nearly any other discipline.

I also contacted the American Council on Education, the American Association of Higher Education, and the Association of American Colleges and Universities to ask whether they had recently sponsored research, papers, or panels on grade inflation. Again, the answer was "no." The U.S. Department of Education had recently begun "profile" surveys of undergraduates[3] and studies of postsecondary transcripts[4] that contained data on student grades, but it issued no warnings on grade inflation. On the contrary, Clifford Adelman, principal author of the transcript studies, claimed that "at most schools there is no grade inflation."[5]

In the summer of 2005, I repeated this survey with minor variations (e.g., I substituted the Council on Higher Education Accreditation for the defunct American Association for Higher Education AAHE) and obtained virtually the same results. The only notable differences were: (1) two executive directors and one associate executive director regional accrediting associations expressed interest in seeing the results of my research; (2) four of the replies mentioned that they required institutions applying for accreditation to provide evidence of assessment standards, such as "clearly stated learning objectives," "established and evaluated learning outcomes," and "student learning outcomes," and that grading data might be supplied voluntarily by an institution to support its case for using such standards. One comment seemed to me especially revealing. It reads:

> Having gone through a very thorough process of Standards revision, I do not recall anyone at any time suggesting that the Commission include in its Standards any items relating to grade distributions. From my experience in accreditation, I think our institutions would find it intrusive and inappropriate for the Commission to make prescriptive statements regarding grade distributions."[6]

Two things are remarkable here. First, in spite of all the controversy over grade inflation in recent years, it is remarkable that an accrediting commission could go "through a very thorough process of Standards revision" without "anyone at any time suggesting" that "items relating to grade distributions" ought to be included. Second, it is noteworthy that this chief officer thinks "our institutions would find it intrusive and inappropriate for the Commission to make prescriptive statements

regarding grade distributions," despite the fact that her commission is not reluctant to make prescriptive statements about assessment of student learning and other faculty responsibilities. Moreover, qualms about *prescribing* good grading practices do not explain why her commission does not ask for institutional data or policies on grade distributions. Are grades so politically charged that data alone are inflammatory? The implicit message here is "don't ask, don't tell."

One thin crack in this wall of silence can be found in the Middle States Commission on Higher Education's *Student Learning Assessment: Options and Resources* (2003). This 100-page document includes a brief section (slightly over one page) entitled "Where Do Grades Fit into the Picture?"[7] Most of this section consists of sensible, if mundane, observations about the relationship between grades and assessment. For example:

> The Commission recognizes that grades are an effective measure of student achievement if there is a demonstrable relationship between the goals of and objectives for student learning and the particular bases (such as assignment and examinations) upon which student achievement is evaluated (Standard 14).[8]

But one short paragraph deals with grade inflation. It reads in its entirety:

> One reason "grade inflation" is seen as a problem is that grades alone cannot be relied on to reflect student performance accurately. One could ask: "Does one grade of 'A' equal another?" If instructors were to match grades explicitly with goals, it would become easier to combat grade inflation, because high grades must reflect high performance in specified areas.[9]

The first sentence contains multiple ambiguities. First, it is not clear whether "seen as a problem" should be taken to mean: (a) "is a problem"; (b) "is believed to be a problem"; or, (c) "is mistakenly believed to be a problem." Second, it is not clear whether "grades alone cannot be relied on to reflect student performance accurately" means: (1) "accurate grading is not possible without good assessment"; (2) "final grades alone cannot provide students with sufficient feedback on their work"; or (3) "grades cannot serve as accurate summative indications of student performance." Third, it is not clear whether the

"reason" referred to is (1), (2), or (3) or the negations of (1), (2), or (3). I am fairly sure, however, that all of the permutations of meaning here are false. I have never come across anyone who claimed either (1), (2), (3), *or their negations* as a "reason for seeing grade inflation as a problem." The simplest reading of this sentence might be, "Grade inflation is a problem because a grade cannot serve as an accurate summative indication of student performance," but this contradicts the commission's assertion that grades appropriately linked to assessment can serve as "an excellent indicator of student learning,"[10] "an effective measure of student achievement,"[11] and so on.

The second sentence also is ambiguous, but the third sentence seems relatively straightforward. It suggests that good assessment practices such as matching grades explicitly with goals and measurable standards will help combat grade inflation. This is a genuinely interesting claim and one that might serve as a core argument for the position that good assessment is the royal road to grade inflation reform. The difficulty is that this core argument is not supported by data. The advocacy of accrediting bodies and national associations for higher education over the past ten years or so has strengthened assessment practices at a great many colleges and universities, and yet during this same period grade inflation has risen to an all-time high. The *National Survey of Student Engagement/Annual Report 2004*, based on a survey of 163,000 randomly selected freshmen and seniors at 472 four-year colleges and universities, found: "About two-fifths of all students reported that they earned mostly A grades, another 41% reported grades of either B or B+, and only 3% of students reported earning mostly Cs or lower."[12] Of course, it is possible that improved assessment has helped in some way to prevent grade inflation from getting even worse, but that is a hypothesis that accrediting bodies and national associations have not sought to test and, to the best of my knowledge, has not been tested independently.

With few exceptions, the organizations that lead American higher education have turned a blind eye to the relationship between the assessment of student learning and the use of grades to report that assessment. As a consequence, they fail to see that the benefits of improved assessment are lost or compromised by grade inflation.

Consider the following parallel (also discussed in Chapter 3) The technicians at Consumer Union are assessment professionals. They are skilled at identifying appropriate goals for a given consumer product,

developing measurable standards for the achievements of those goals, and testing the effectiveness (i.e., outcomes) of various models in meeting those standards. Self-propelled lawn mowers, for example, are tested against standards for evenness, mulching, bagging, side discharging, handling, and ease of use (with a separate report on brand repair history). The results of these tests are reported in *Consumer Reports*, where models are ranked according to their overall scores, and each model is graded "excellent," "very good," "good," "fair," or "poor." A reader of the report can see at a glance the relative strengths and weaknesses of each model as well as overall ranking. Now imagine that instead of providing information of this kind, *Consumer Reports* reported its findings by assigning a grade of "excellent" to 40 percent of the models tested, a grade of "good' or lower to 3 percent of the models, and a grade of "very good" to the remainder. Were this to happen, no one would be surprised to see the readership of *Consumer Reports* plummet. Fortunately, the managers and technicians of Consumers Union recognize that they need to set standards high enough to differentiate the performance of the models they test and to report those differences to their readers.

Old habits die hard, and transcripts of college grades often are sent out as a matter of course, but there is evidence that their readership is declining. Rosovsky and Hartley cite a survey of Human Resources Officers (HRO) from Fortune 500 companies in 1978, 1985, and 1995 that "found that the percentage of HROs who agreed that transcripts of college grades ought to be included with an applicant's resume fell from 37.5 percent to 20 percent."[13] This is truly unfortunate, for it suggests that employers are paying less attention to the judgments of teachers and more to standardized texts and institutional pedigrees. The greater misfortune, however, is the effect of grade inflation on the motivation of students. At the high end, diligent students are not being motivated to stretch themselves beyond the comfortable range in which A is more or less guaranteed. As one Princeton undergraduate explained:

> I have had the following experience multiple times in my Princeton career: I don't get a chance to study for a test as much as I'd like/work on a paper for as long as I'd like. I know that the product I turn in is a far cry from my best work. I get the paper back with an A on it. Although I fight against the instinct, I feel much less motivated to work hard to learn in that class in the future. The result? I learn less.[14]

On the low end, lazy students can earn passing grades (often in the B range) simply by going to class and turning in assignments. The *National Survey of Student Engagement/Annual Report 2005* found: "By their own admission, three of ten first-year students do just enough academic work to get by."[15]

The depth of the motivation gap in American higher education and the failure of improved assessment practices to lessen this gap are revealed by the number of hours that students spend on academic work outside of class. Although it is widely accepted that students should spend at least two hours outside of class for each hour in class (about thirty hours a week for a full-time student), few students come close to this minimum. Since its inception in 2000 (a pilot was conducted in 1999), the *National Survey of Student Engagement* has asked randomly selected freshmen and seniors at four-year colleges and universities to indicate the number of hours they spend "preparing for class (studying, reading, writing, rehearsing, and other activities related to your academic program)." In 2000, only 12.1 percent of first-year students and 14.5 percent of seniors said that they spent more than twenty-six hours a week preparing for class. In 2005, only 8 percent of first-year students and 11 percent of seniors said they spent more than twenty-six hours a week preparing for class.

If widespread reform is to be achieved, then policymakers in higher education must stop ignoring grade inflation and start sponsoring more research on grading practices, more discussion of consequences, and more leadership for change. Given national and regional backing, both institutions and individuals will be in a stronger position to create models for reform. National leadership is essential because the intelligibility of grades depends on nationally accepted conventions.

INSTITUTIONAL REFORM

At the institutional level, reform can deal with grades themselves or with factors that contribute to grade inflation. Among the latter are efforts to modify the design of student evaluations of teaching (SET) and their application to personnel decisions such as tenure and promotion. A simple step is to include grade distribution data in applications for tenure, promotion, and so on and to weight teaching valuations according to the rigor of an applicant's grading. In my experience as a dean and department chair, a tough grader who gets consistently

high evaluations is likely to be an outstanding teacher. Another approach is to adjust SET scores upward or downward to reflect grading rigor or leniency. This approach has been adopted by the University of Washington.[16] Reforming SET instruments, interpretation procedures, and applications in personnel decisions is worth doing for many reasons, but it remains to be seen whether it will lead of its own accord to a reduction in grade inflation. The most I think we can count on is that it will reduce an unhealthy pressure within the academy that helps nurture grade inflation.

Other small but worthwhile steps include limiting or eliminating pass/fail options, course withdrawals, no-penalty Fs, substitution of grades from repeated course grades, and the inclusion of transfer grades in cumulative GPA and honors calculations. Harvard's faculty made news in 2002 by voting to limit Latin honors to 60 percent of its graduates of whom no more than 3 percent can be *summas*.[17] Sewanee: The University of the South "generally" does not calculate grades from transfer courses in a student's GPA and restricts attempts to improve grades by repeating courses by stipulating: "If, and only if, the earlier grade was lower than C– will both grades be calculated into the cumulative grade average."[18]

Among the ways to deal directly with assigned grades are three principal strategies for reform: (1) adopting caps that limit the percent of grades that can be awarded; (2) adjusting upward the value of submitted grades for a class when the class average is lower than a preferred average and downward when the class average is higher than a preferred average; (3) adding information on transcripts that clarifies or supplements submitted grades.

The adoption of grade caps is the most straightforward way to change grade distributions. It is both direct and transparent. Caps could be used to mandate that the grades submitted for every class (or every department) conform to a strict (or modified) normal curve. The chief drawback with caps is that they may be seen by faculty members as an infringement on academic freedom. Faculty accustomed to grading as they please may find it repugnant to curb their judgment in order to conform to university policy.

Remarkably, the Princeton University faculty voted on April 26, 2004, to adopt a quota on undergraduate A-level grades. By a vote of 156 to 84 (of those present and voting), they approved a proposal, described as "a social compact,"[19] requiring each department or program

to assume responsibility for meeting the institutional expectations that in "undergraduate courses As (A+, A, A-) shall account for less than 35 percent of the grades given in any department or program in any given year" and in "junior and senior independent work As (A+, A, A–) shall account for less than 55% percent of the grades given in any department in any given year."[20]

Viewed from the outside, the adoption of this policy appeared to be a swift and daring decision by a traditional research faculty. In fact, it was the culmination of eight years of study, debate, and experimentation and owed a good deal to the patient and persistent leadership of Nancy Weiss Malkiel, Dean of the College. In February 1998, Dean Malkiel sent the faculty the results of a two-year study of grading patterns at Princeton from 1973 to 1997, compiled by the Faculty Committee on Examinations and Standing (a committee that she chaired). Its purpose was to stimulate reflection, encourage change, and solicit feedback. The following September, the committee sent a progress report to the faculty with several proposals for grading reform. The upshot was the development of a *Guide to Good Grading Practices*; annual distribution of extensive grading data; asking faculty to complete a grid on each grading sheet that counts final grades in each category; and adoption of a policy that makes an A+ equivalent to an A (4.0 instead of 4.3) for purposes of grade calculation. By 2003, it was clear that these steps had not succeeded in stemming the rise of GPAs or the compression of grades to the upper end of the grading scale. In the period 2001–2002, a senior with a straight-B average (3.0) ranked 923 out of a graduating class of 1,079, and a student with a straight-C average (2.0) ranked 1,078. Faced with these disappointing results, a majority of department chairs called for collective action on a simple institution-wide grading policy. The committee responded with a proposal for capping As at 40 percent. This was reduced to 35 percent in the final version.

Some details of Princeton's policy bear further examination. The first is the capping of As overall to 35 percent and the capping of As for junior- and senior-level independent work to 55 percent. Why these percentages? The supplementary attachment "Grading Questions and Answers" mentions the following: (1) "the limits needed to be achievable in the Princeton context"; (2) "the difference between the proposed limits and the current practice needed to be large enough to be meaningful"; (3) "35/55 . . . resembles grading patterns at Princeton . . . in the

period F87–S92 . . . describes . . . current grading patterns in some departments"; (4) "if we really mean to make a difference in tackling grade inflation, we should take a bigger rather than a smaller step." Item 3 is stated more emphatically by Dean Malkiel in "Explaining Princeton's New Grading Policy," in *Princeton Parents News* (Summer 2004). She says:

> From 1973 to 1987, As accounted for just under a third of the grades we awarded in undergraduate courses; from 1987 to 1992, the figure was between 36 and 37 percent. It's only been in the last decade that the percentage of As has spiked upward (today, it's over 47% percent). We believe that our historic patterns better reflect an honest appraisal of the distribution of student academic performance.[21]

The difficulty with this appeal to "historic patterns" is that the benchmark period (1973–1992) mentioned here was one in which grade inflation had already reached a historic high. Why would Princeton not have chosen the much longer period of grade distribution stability that preceded the mid-1960s, a period in which As accounted for a smaller percentage of all grades? I asked Dean Malkiel this question. She replied that until 1969, Princeton graded on a 7-point scale (1 to 7), and that good data on grading were not collected until 1973. She also indicated that the chosen caps represented what was politically feasible.

Another key detail is the responsibility of departments and programs. Although the "expectations" (i.e., quotas) are not optional, "it shall be left up to each department or program to determine how best to meet these expectations."[22] While this arrangement could lead to political battles within departments and programs, it has the distinct advantage of giving faculty members in each department or program the responsibility for working out the details of distribution and implementation. It also goes a long way toward addressing the serious problem of grading inequities among disciplines.

Oversight of departmental progress in meeting these expectations is lodged in the Faculty Committee on Examinations and Standing and a Grading Committee made up of one department chair from each of the six divisions plus the Dean of the Faculty, the Dean of the College, and the Registrar. Each year the former committee reports "to the faculty on the grading record of the previous year," but "the

standard by which the grading record of a department or program will be evaluated will be the percentage of As given over the previous three years."[23] In departments that fail to meet grading expectations, the Committee on Grading shall advise the other committee "on an appropriate strategy to assure adherence in the future."[24]

An important insight into the thinking behind this policy can be found in a supporting document from the Office of the Dean of the College. The concluding paragraph of that document reads:

> Princeton enrolls a select group of unusually accomplished—indeed, increasingly accomplished—students whose credentials and achievements place them in the front rank of undergraduates in all American colleges and universities. The new grading policy reflects the commitment of the Princeton faculty to hold these students to the highest standards and to make very careful distinctions in evaluating their work. Princeton grades should be understood as rigorous markers of academic performance in an extremely challenging academic program of undergraduate study.[25]

What is crucial here is the principle that the standards for grading and assessment at Princeton should be commensurate with the capabilities of current Princeton undergraduates. This is very different from the common practice of setting "high" standards that are not defined with reference to any particular set of students or with reference to recollections of former students or some imagined national average.

As one might expect, A-level grades are defined in this policy in fairly rigorous language. An A+ means to "significantly exceed the highest expectations for undergraduate work." An A "meets the highest standards for the assignment or course." An A– means "meets very high standards for the assignment or course." Yet grades from B to C– are defined in surprisingly undemanding terms. A "B" grade, for example, means "meets most of the standards for the assignment or course." A "C" grade means "meets some of the basic standards for the assignment or course." One is tempted to ask why a student who succeeds only in meeting "some of the basic standards for a course" should pass the course at all. A good deal of careful thought seems to have gone into crafting that portion of Princeton's grading policy that deals with As, but the same level of care does not appear to have been given to other grades. "Grading Questions and Answers" states that

"this plan is premised on the assumption that if we control A grades, other grades will fall into line." This seems to me overly optimistic.

In the period 2004–2005, the first year under the new policy, A-level grades dropped from 46 percent in 2003–2004 to 40.9 percent. The sharpest drop occurred in humanities departments, where A-level grades went from 56.2 percent to 45.5 percent. There were more B+, B, and B– grades, but only a tiny increase in C grades. Strategies for reducing As varied among departments. The Economics Department agreed on specific targets for types and levels of courses. The English Department left the implementation of the policy to individual faculty.

Although a goal of 35 percent As is modest in comparison to the 15–20 percent that prevailed until the mid-1960s, Princeton's new policy is, nonetheless, a bold and laudable initiative that may signal a new willingness among faculty to grapple with grade inflation through consensus. On April 14, 2004, Wellesley College's Academic Council approved a resolution on grading that included the provision "The average grade in 100- and 200-level courses should be no higher than B+ (3.33)."[26] MBA classes of 20 or more at the Wharton School of Business cannot excede a 3.33 average GPA.

What about other options? The adjustment or indexing of assigned grades also is a direct approach to managing grading patterns, but it is likely to be less transparent. It requires that a formula (possibly of some complexity) be adopted for adjusting the value of submitted grades and may result in guesswork on the part of students and faculty about strategies for maximizing grade value from semester to semester. One of the simplest adjustment schemes was proposed by Noel deNevers in 1984—subtract the class average from each submitted grade and then add whatever preferred average grade the college has established.[27] Thus if the submitted grade for a student is 3.5, the class average is 3.4, and the college's preferred average grade is 3.0, then the adjusted grade would be 3.1. (This adjusted grade could replace the submitted grade entirely, be used for the calculation for GPA, or simply appear alongside the submitted grade.) Other adjustment schemes, such as that proposed by Larkey and Caulkin, use a linear regression model to compensate for actual disparities in the grading leniency or severity of different departments and programs.[28]

Valen Johnson has pointed out that these relatively simple models can create penalties for students who score top grades in leniently graded classes. Consider two students, Tom and Jerry, who have taken

identical classes every year and scored an A+ in each course. In their senior year, Tom takes an extra class in drama and scores an A+, while Jerry does not take an extra class. Because the drama class is graded very leniently, Tom receives an adjusted grade of A– and ends up with a lower GPA than Jerry. To overcome this unfair penalty and more subtle problems, Johnson has proposed "a more sophisticated, nonlinear statistical model called a multirater-ordinal probit model."[29] Although Johnson's model succeeds in compensating for grading inequities among disciplines without producing unfair penalties, mathematically unsophisticated faculty, students, and parents are likely to regard his model as a black box that does unintelligible things to submitted grades and, therefore, to resist his solution. Whether resistance of this kind can be overcome is an empirical question, but I suspect it cannot.

Adding information on transcripts to clarify or supplement submitted grades is likely to be more palatable to faculty than proposals that restrict or adjust submitted grades. For example, undergraduate transcripts and student grade reports at Dartmouth College indicate the median grade earned in a class and class enrollment. Columbia University reports the percentage of A-range grades for classes, colloquia, and seminars over a certain size on undergraduate transcripts at Columbia College and the School of General Studies.

The challenge is to add precise information that effectively compensates for grade conflation and can be readily understood by readers of transcripts. In "Grade Inflation: Metaphor and Reality" (2004), Mary Biggs and I propose a ranking model that meets this challenge.[30] We believe an F should be given to students if and only if they have failed to achieve the learning objectives of a course or failed to complete the assignments required to demonstrate achievement of those learning objectives. However, we recommend that all students who receive passing grades should be ranked. Each teacher would submit a ranked student list, with rank-sharing either disallowed or designed to approximate a normal curve. For example, there might be four ranks for passing students: top 15%, next 25%, middle 40%, and bottom 20%. The grades and ranks for each course and a running CRA (class rank average) would be displayed on the transcript, along with a boilerplate explanation. The aim of our recommendation is to require teachers to use their good judgment, professional skills, and unique role as observers of student work to provide precise information about relative performance. Although we prefer that grades be used to achieve this

aim, we contend that ranking would provide comparable information and help pave the way for thoroughgoing grade reform. Required ranking, whether *seriatim* or on a normal curve, would force teachers to confront the disparities of grade conflation every semester and could prepare them psychologically for a return to normal curve grading.

As a weaker expedient for campuses that are unwilling to adopt ranking, we see some value in listing the mean grade (or similar data) for each course on a student's transcript, as Dartmouth has done. The advantage here is ease of implementation: faculty grade as they please, and the registrar's office provides compensating information. While this is clearly a step in the right direction, it is less precise than ranking, and its precision diminishes with grading leniency. The college where I teach, the College of New Jersey, has a well-respected School of Education. In the spring of 2005, the percentage of A-range grades (A or A–) awarded by its Department of Languages and Communication Sciences (not our Modern Language Department) was 90 percent; the percentage awarded by Elementary/Early Childhood Education was 85 percent. Ranking could tell me that Jane, who earned an A in Elementary Education 101, was at the very top of her class, while John, who also earned an A, was near the middle, but posting a mean class grade next to their As will not differentiate their achievements. Another drawback of this weaker expedient is that it makes good faith in grading an administrative function rather than a faculty responsibility.

All things considered, I believe the best reform is the simplest and most transparent. Earlier in this chapter I suggested that most of the problems that have accumulated from four decades of grade inflation could probably be corrected by the adoption of just three principles:

1. No more than 20 percent of undergraduate grades at the A level (i.e., A+, A, or A–, or their equivalents)
2. No more than 50 percent of undergraduate grades at the A and B levels combined.
3. The grade "F" is used to indicate failure to meet the standards for a course and not just failure to attend classes or complete assignments.

This is a controversial prescription, and some elaboration is in order. A 20 percent cap on As may seem too generous to some reformers and too stingy to others. I have suggested 20 percent in hopes of

making the cap both tough and attainable. The absence of any stipulation regarding Ds may seem like an oversight, but it is not. Although I am strongly in favor of using D's wherever appropriate, I am reluctant to defend a fixed distribution of D's. My reluctance is based on the conviction that even in the heyday of normal curve grading D was understood to be an indicator of passable but substandard achievement. In other words, I believe that D, like F, has always carried a connotation of deficiency distinguishable from its role as a marker of relatively low class rank. Just as I am unwilling to recommend that some minimum percentage of students must fail a course and some minimum receive As, I am reluctant to prescribe that some minimum who pass the course must be identified as deficient. In this respect, I am unsympathetic to true normal curve grading. I also am unsympathetic to the practice of distributing grades on a normal curve, regardless of how successful or unsuccessful students are in meeting the standards of a given course. I think faculty members ought to adjust and readjust assessment goals and standards in order to bring into line the grades their students earn with institutional maxima. I believe the adoption of these principles, coupled with implementation procedures modeled on Princeton's new grading policy, particularly the assignment of responsibility for compliance to individual departments and programs, holds the greatest promise for thoroughgoing reform.

The essential thing is to act and be heard. Given the stunning silence of accrediting bodies and national associations on the problem of grade inflation, individual faculty members, departments, and institutions must take the initiative. All of the steps previously mentioned are worth considering, but it is crucial that whatever steps are taken be taken in a very public way. Our colleagues, administrators, professional associations, and accrediting bodies and the public at large need to be told that grade inflation is real, and that it is not okay. They need to hear that grading in ways that deny students information that could be useful for choosing majors and careers and deprive employers and graduate programs of a counterbalance to standardized test scores or reliance on the criterion of institutional prestige is not a legitimate exercise of academic freedom. Grade reform is not, as Alfie Kohn has suggested, a return to barbaric times when we were "trying to make good students look bad."[31] In 1960, a B was a good grade, a mark of superior performance, and C was a decent grade. Grade reform is

not about being tough but about being honest. Our students have a right to know how well they are doing, and we have a duty to tell them.

NOTES

1. Henry Rosovsky and Matthew Hartley, "Evaluation and the Academy: Are We Doing the Right Thing?: Grade Inflation and Letters of Recommendation," Occasional Paper (Cambridge, MA: American Academy of Arts and Sciences, 2002), 14.

2. The 2005 *High School Survey of Student Engagement (HSSE)*, which collected data from 180,000 students from 167 high schools in twenty-eight states, found: "89% reported grades of B or better with 44% reporting either A or A–" and "Two-thirds of those who study three or fewer hours a week reported receiving mostly A and B grades." Center for Postsecondary Research, *National Survey of Student Engagement/Annual Report 2005* (Bloomington: Indiana University School of Education, 2005), 30–31.

3. Laura Horn and Jennifer Berktold, *Profile of Undergraduates in U.S. Postsecondary* Institutions, 1995–1996 (Washington, DC: U.S. Department of Education, National Center for Education Statistics, May 1998).

4. Clifford Adelman, *The New College Course Map and Transcript Files: Changes in Course-Taking and Achievement, 1972–1993* (Washington, DC: U.S. Department of Education, 1995); Clifford Adelman, Bruce Daniel, and Ilona Berkovits, *Postsecondary Attainment, Attendance, Curriculum, and Performance: Selected Results from the NELS: 88/200 Postsecondary Education Transcript Study* (Washington, DC: U.S. Department of Education, National Center for Education Statistics, September 2003.

5. Clifford Adelman, "A's Aren't That Easy," *New York Times* (May 17, 1995), p. A19.

6. As a courtesy to association officers who responded to my survey with comments, I am not citing officers or associations by name.

7. Middle States Commission on Higher Education, *Student Learning Assessment: Options and Resources* (2003): 36–37.

8. Ibid., 37.

9. Ibid.

10. Ibid., 36.

11. Ibid., 37.

12. Center for Postsecondary Research, *National Survey of Student Engagement/Annual Report 2004* (Bloomington: Indiana University School of Education, 2004), 13.

13. Rosovsky and Hartley, "Evaluation and the Academy," 12. The article they cite is Nelda Spinks and Barron Wells, "Trends in the Employment Process: Resumes and Job Application Letters," *Career Development International* 1999.

14. Quoted in Nancy Weiss Malkiel, "Explaining Princeton's New Grading Policy," *Princeton Parents News* (Summer 2004).

15. Center for Postsecondary Research, *National Survey of Student Engagement/ Annual Report 2005*, 12.

16. At the University of Washington, median scores for some of the questions on student evaluations are adjusted on the basis of students' reasons for enrollment, class size, and student responses to the following question: "Relative to other college courses you have taken, do you expect your grade in this class to be: Much Higher, Average, Much Lower" (answered on a 7-point scale). "Adjusted Medians," Office of Educational Assessment, University of Washington, http://www.washington.edu/oea/services/course_eval/uw_seattle/adjusted_medians.html. As a result, faculty are given slightly higher scores if students *expect* their grades to be lower than average and lower scores if students expect their grades to be higher than average. I asked the university's Office of Educational Assessment for information on faculty reactions to this practice and longterm consequences, but it had data on neither.

17. A. Hartocollis, "Harvard Faculty Vote to Put the Excellence Back in the A," *New York Times* (May 22, 2002), p. A20.

18. "Academic Policy: Student Classification, Progress, and Status," e-Sewanee: The University of the South's home on the Web, http://www.sewaneetoday.sewanee.edu/catalog/details/585.html.

19. "Report to Faculty on Grading Proposals" (April 6, 2004). Internal document of Princeton University. Quoted phrase is from the second section of this document "Grading Questions and Answers," which is separately paginated, p. 2, http://www.princeton.edu/~odoc/grading_proposals/.

20. "Princeton University Grading Policies," Approved by the University Faculty, April 26, 2004, 1. Posted on the Princeton University Web site at http://www.princeton.edu/~odoc/grading_policies.htm. Also posted on this Web site are "Grading Definitions"; "Princeton University Grading Policies in Undergraduate and Independent Work," from the Office of the Dean; "From Grading Questions and Answers"; Nancy Weiss Malkiel, "Explaining Princeton's New Grading Policy" *Princeton Parents News* (Summer 2004).

21. Weiss Malkiel, "Explaining Princeton's New Grading Policy," 10.

22. "Princeton University Grading Policies," 1.

23. Ibid.

24. Ibid.

25. "Princeton University Grading Policies in Undergraduate and Independent Work," from the Office of the Dean, 1; http://www.princeton.edu/~odoc/grading_statement.htm.

26. "Frequently Asked Questions about Wellesley's New Grading Policy," Wellesley College Web site, Division of Student Life, Policies and Legislation, 1, http://www.wellesley.edu/DeanStudent/gradingfaq.html.

27. Noel deNevers, "An Engineering Solution to Grade Inflation," *Engineering Education* 74 (April 1984): 661–63.

28. P. Larkey and J. Caulkin, "Incentives to Fail," Technical Report, 92-51 (Heinz School of Public Policy and Management, Carnegie Mellon University, 1992). See discussion of grading equity issues in Valen E. Johnson, *Grade Inflation: A Crisis in College Education* (New York: Springer-Verlag, 2003), 196–232.

29. Johnson, *Grade Inflation*, 219.

30. Richard Kamber and Mary Biggs, "Grade Inflation: Metaphor and Reality," *The Journal of Education* 184:1 (2004): 31–37.

31. Alfie Kohn, "The Dangerous Myth of Grade Inflation," *The Chronicle of Higher Education*, November 8, 2002, Section 2, B8.

This page intentionally left blank.

CHAPTER TEN

Grade Distortion, Bureaucracy, and Obfuscation at the University of Alabama

DAVID T. BEITO AND CHARLES W. NUCKOLLS

Accountability and transparency have recently become magic words for Americans, and for good reason. Bitter experiences with corporate scandals have led to public demands for a full accounting of such statistical indicators as assets, value, and liabilities. In higher education, probably the most accurate measure of accountability is student grading. Without grade distribution data, taxpayers, faculty, and students lack the necessary tools to measure success or failure. For this reason, Americans have a right to expect full transparency in grading from colleges and universities. Much to our dismay, we have come to the conclusion that this is not true at the school where we teach, the University of Alabama (UA). As of this writing, the University, through its Office of Institutional Research (OIR), is systematically denying requests from the faculty and the public for data on grade distribution.

This closed-door policy began under the administration of Robert E. Witt, who took office as the president of the University in March 2003. Witt never publicly (or, as far as we know, privately) explained his reasons for deciding to restrict access. We suspect, however, that at least in part he acted in response to our earlier study that had found significant grade inflation and disparities. We know from personal experience that these revelations caused great embarrassment for University of Alabama administrators, who put unusual stress on the importance of "image."[1]

When we began our study of grade distortion, we had never imagined where it would lead. As a historian and an anthropologist, our previous research pointed in entirely different directions. Nevertheless, like many instructors in the classroom, we had a personal and professional

interest in the question of academic standards. This concern motivated us, along with other faculty, to form the Alabama Scholars Association in 2001. It also prompted us to more closely examine grading practices at our own institution. We had read stories about grade inflation at Ivy League schools and wondered if this same problem existed at our own school.

Our request to the OIR asked for data on the overall percentage of As for Fall 2000, Spring 2001, Fall 2001, and Spring 2000 for each college and for the university, as well as for all 100- and 200-level courses in the College of Arts and Sciences (A&S). A&S is the largest college at the university and includes such diverse departments as history, anthropology, biology, and mathematics. Courses at the 100 and 200 level are gateway classes for freshmen and sophomores. Because they are of an introductory nature, they winnow out students before they can proceed to more advanced courses. Thus the percentage of As in gateway courses is generally, or should be generally, lower than in 300- to 500-level courses.

As we awaited an answer from the OIR, we did not expect to encounter much difficulty in getting it filled. In the recent past, the office had routinely complied with faculty requests. During the 1990s, we had easily secured grade distribution data (no questions asked). Moreover, it was common knowledge that colleges and universities throughout the United States were showing a new spirit of candor about grade inflation. We hoped that administrators at the University of Alabama shared in this spirit, at least in a small way. Some schools, such as Indiana University, had gone the extra mile and posted on the Web grade distribution statistics of each instructor, department, and college.[2]

From the start, however, we encountered great difficulty in getting the requested information from the OIR. It was the beginning of a pattern. The OIR failed to promptly answer our e-mails and phone calls and engaged in other delaying tactics. Apparently, somebody at either the OIR or central administration had decided to pull back from the previous policy of cooperation and free access. After we unilaterally announced our intention to come in and copy the data (at our own expense), officials at OIR reluctantly agreed. Since at least the early 1970s, the OIR had produced printouts at the end of each semester and put them in bound volumes. These showed grade distribution data for every department and college. The OIR's printouts

for the 1970s through early 1980s are freely available to researchers at the Hoole Library (special collections) at the university.

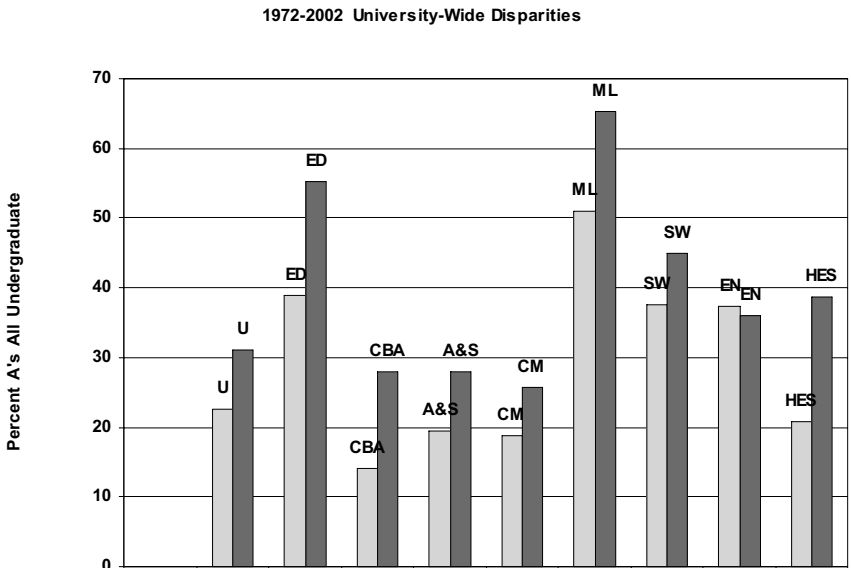
Fortunately, a 1976 study of grade distribution at the University of Alabama enabled us to make a thirty-year comparison. Published by the OIR (then called the Office of Institutional Analysis), it included data on the percentage of As at both the college and university levels over several semesters. Apparently the 1976 study had pretty much gathered dust at the Hoole Library until we made use of it. To determine a final figure, we averaged out the percentage of As over four semesters: Fall 1972, Spring 1973, Fall 1973, and Spring 1974.

Our comparison of the percentage of As from 1972 to 1974 to those from 2000 to 2002 showed significant grade inflation. We also found pronounced disparities between units for the university in otherwise comparable 100- and 200-level gateway courses. Charles Nuckolls coined a new term to encompass these interrelated problems—grade distortion. Grade inflation, the first component of grade distortion, describes a rise in letter grades awarded to students over a defined period. The level of grade disparity can be measured by calculating the differences between units internal to the university (colleges or departments) in the percentage of high letter grades awarded to students in a defined period.

The 1976 study revealed that grade inflation was already underway. Between 1972 and 1974, the average percentage of As for undergraduates at the university was 22.6 percent. The Office of Institutional Analysis considered this high and warned that "the percentage of As and Is awarded has been steadily increasing," especially among undergraduates. In our experience, this willingness to candidly grapple with the problem of grade inflation, as well as to publish less than flattering distribution statistics, provided a dramatic contrast to the OIR's closed-door policy in the 2000s.[3]

The 1976 report's warnings about grade inflation fell on deaf ears, and grade inflation continued to accelerate during the next two decades (see Table 10.1). From 2000 to 2002, the percentage of As in all undergraduate courses was 31.1 percent. This represented a significant increase of 37.6 percent from 1972 to 1974. One of the worst offenders in both periods was the College of Education. From 2000 to 2002, As constituted 55 percent of all undergraduate grades in that college.[4]

What has caused grade inflation at the University of Alabama? In 1996, the OIR concluded (in a report not released to the public) that



	U	ED	CBA	A&S	CM	ML	SW	EN	HES
1972-74	22.6	39	14	19.4	18.7	50.9	37.5	37.3	20.7
2000-02	31.1	55.3	27.9	27.9	25.8	65.4	45	36	38.6

Note: It is not clear that the College of ML, "Military Science," is the direct successor to ROTC Army in the earlier period. However, we have treated it as such for this anaylsis. HES, "Human Environmental Science," is the successor to Home Economics.

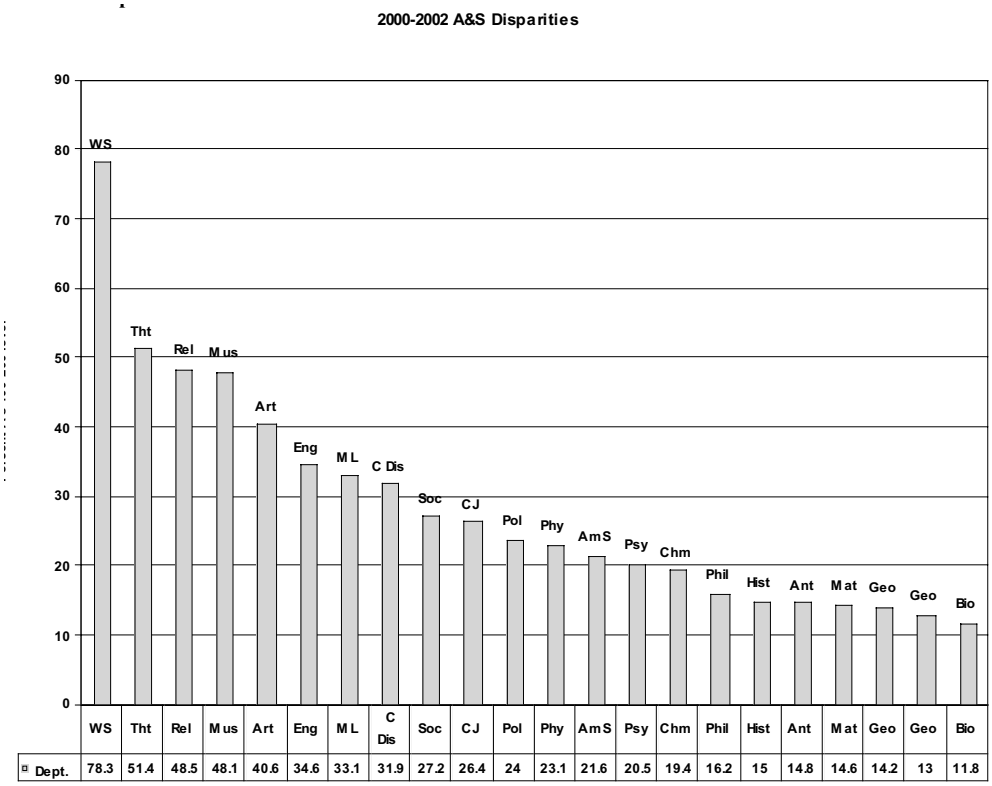
it was due to "admission of better prepared high school graduates." We do not find this explanation convincing. From 1972 to 2001, the average ACT scores for entering freshmen increased relatively little (from 22.9 to 24.5), an amount difficult to reconcile with the 37.6 percent increase in undergraduate As over the same period. According to Bob Ziomek, director of ACT program evaluation, "The ACT average doesn't explain the whopping increase in As being awarded. ACT scores are fairly flat while the number of As and Bs being awarded are out of sight." Even if the claims of improved student quality are true (and we are dubious that they are), it does not necessarily follow that grade inflation is justified. We share Harvey Mansfield's view that if students "are in some measures better, the proper response is to raise our standards and demand more of our students. So when buying a car, would you be satisfied with one that was as good as they used to be."[5]

In studying the second, and in our view, more serious, component of grade distortion, grade disparity (see Table 10.2), we limited our focus to 100- to 200-level gateway courses in the College of Arts and Sciences from 2000 to 2002. The contrasts were striking. The most inflated department in A&S was Women's Studies. Between 2000 and 2002, the average percentage of As in that department averaged an almost unbelievable 78.1 percent. Other highly inflated departments were Theater/Dance (51.4), Religious Studies (48.5), and Music (48.1). The five least inflated departments were Biological Sciences (11.1), Geography (13), Geological Sciences (14.2), Math (14.6), and Anthropology (14.8).[6]

Although our departments (Anthropology and History) fell in the 15 percent or lower range, we make no claim of innocence. The differences are relative at best. Grade disparities exist within nearly every department on campus. The Blount Undergraduate Initiative is a case in point. Students in the initiative, an interdisciplinary residential program on campus, take a common set of courses from a select group of instructors.

From 2000 to 2002, the percentage of As varied dramatically in Blount courses, which randomly assigned students and gave the same readings. In the spring semester of 2002 (which is typical of previous semesters), one instructor gave 81.7% As in such a course, while another gave 37.4. The other five awarded 53.7%, 46.6%, 38.4%, and 46%. In the fall of 2000, the numbers are nearly as disparate: 55.1%, 38.6%, 56.2%, 47.0%, 47.4%, and 58.3%. In the spring of 2001, the class averages for number of As awarded were 64.5%, 18.7%, 56.2%, 69.8%, 83.2%, 18.7%, and 71.5%.[7]

Grade disparity of this extreme nature serves to undermine educational quality and standards. It also shortchanges the best and hardest working students. When grade disparity is rife, as it is at the University of Alabama, the overall grade point average can no longer be said to adequately reflect comparative abilities. The grade of the A student in the course that demands little effort is placed on an equal plane with that of the student who has to struggle to earn the same grade in a more difficult course. The system creates perverse incentives for students to "shop around" for professors who have reputations for giving "easy As" and serves to degrade the efforts of those students who might otherwise take "harder" courses. Under such a system, the student transcript loses its value as a source of information for



Note:

WS	Women's Studies
Tht	Theater and Dance
Rel	Religious Studies
Mus	Music
Art	Art History
Eng	English
ML	Modern Languages
C Dis	Communication Disorders
Soc	Sociology
CJ	Criminal Justice
Pol	Political Science
Phy	Physics & Astronomy
AmS	American Studies
Psy	Psychology
Chm	Chemistry
Phil	Philosophy
Hist	History
Ant	Anthropology
Mat	Mathematics
Geo	Geological Sciences
Geo	Geography
Bio	Biology

potential employers who need to judge the comparative qualifications of UA graduates.

Once we completed our original study showing grade distortion, we e-mailed the results to department chairs, deans, and other administrators to get their reaction. Few bothered to respond. We also summarized our results in a Power Point presentation at a special meeting with J. Barry Mason, the interim president of the University of Alabama. President Mason said that our study raised important issues that deserved serious consideration by both faculty and administrators. He urged us to contact him if we had any more difficulties with the OIR, and he pledged to intervene to help us. Later, President Mason sent out a general e-mail to the faculty praising our report: "It is a delight to see the faculty tackling this issue in such a forthright manner as we continue to make sure this University is known for its excellence in academics."[8]

We shifted our attention to the Faculty Senate, hoping to get it to implement reform. Now that the problem of grade distortion at the university was public knowledge, we weighed the advantages and disadvantages of several possible suggestions for reform. From the outset, we agreed that the best approach was to promote transparency in grading rather than interfere with faculty discretion. Along with Professor James Otteson, we sponsored a resolution in the Faculty Senate to adopt a system used at Dartmouth College. At Dartmouth, all student transcripts not only include the grade for the class but also the average grade for all students enrolled in the class. This approach does not restrict faculty freedom in grading practices but enables prospective employers and graduate schools to get a better idea of whether that A– is to be admired or ignored. While it is not clear if the Dartmouth system reduces grade inflation, it has the virtue of promoting truth in advertising.[9]

Throughout late 2002 and early 2003, we worked to build public support for this proposal. Articles appeared in the *Tuscaloosa News* and the *Birmingham News*, and the Alabama Scholars Association sent a summary of our study to members of the state legislature. The mailing suggested that the state of Alabama undertake a "grade distortion" audit to determine the extent of the problem in K–12 and in colleges and universities.[10]

Although the reaction from the press and public was generally positive, most members of the Faculty Senate did not indicate any concern. The final vote of the Senate in February 2003 rejected even

our modest proposal to list the class grade point average on each transcript. Not surprisingly, those senators from divisions that had high percentages of As, including Education, Criminal Justice, and Social Work, proved especially hostile. One professor of the College of Education even proudly confessed in an open Senate meeting that he was a "proud grade inflator," and that his main goal was to introduce students to issues of race, class, and gender. We asked our opponents (several of whom admitted that grade inflation was a serious problem) to suggest alternatives, but none expressed any interest in doing so.[11]

Our experiences are consistent with those described by Valen E. Johnson in his recent book on grade inflation. He observes that faculty senates "are populated by the very individuals who benefit most from inequitable grading policies. Or they represent departments that do. Appropriate strategies for reform thus depend on the integrity of the individuals who compose these bodies." Since the defeat of our proposal, the Faculty Senate has shown no indication of wanting to revisit the issue.[12]

Despite this defeat, we had at least some consolation, or so we thought. We told ourselves that we could continue to publicize the problem for future updates and thus create a foundation for eventual reform. To this end, we embarked on a follow-up to our original study. In 2003, we asked the OIR for all the grade distribution data for Fall 2002 and Spring 2003. We wanted to find out if our study had prompted any change in faculty grading behavior. This time the OIR refused to cooperate in no uncertain terms. We had a long and frustrating correspondence with William Fendley, the head of the OIR. We offered to copy the data ourselves (as we had before) from the bound volumes at the OIR.

Although we found it hard to believe, Fendley told us that the OIR had stopped printing out this data because of "workload" and "budget" issues. This was apparently the first time in all the years that Fendley had compiled the data that this problem had come up. The contrast with the more cooperative behavior of the University of Alabama at Birmingham was revealing. During this period, it fully complied with a similar request from the Alabama Scholars Association. It was clear to us that the OIR found our previous study embarrassing and wished to prevent any future revelation of the facts. At this point, we wrote to President Witt, asking him to intervene. In contrast to the open and cooperative approach of his predecessor, J. Barry Mason, Witt never

responded to our letter or e-mails. We do not expect any change as long as he is president.[13]

What lessons have we learned? First, while many faculty members at the University of Alabama realize that grade inflation is a problem that should be addressed, a majority of those in power to implement reform, either because of a vested interest or philosophical considerations, prefer the status quo to any action at all. It also is quite possible that administrators favor an inflationary environment because high grades (in their view) serve to attract and retain fee-paying students. Our students generally do not know that the value of their transcript goes down year after year and after as the problem of grade inflation worsens. When they find out, usually years after graduation, it is too late.

Any reform, in our view, must come from pressure exerted by an informed public or by public officials. The most important lesson, however, is the necessity to have full and accurate reporting. This reporting will never occur unless parents, taxpayers, and public officials hold their universities and colleges to the same standards of transparency as they now require of corporations and government.

NOTES

1. The preoccupation with "image" of University of Alabama administrators is not new. Since 1981, for example, the media relations office has conducted an "ongoing image survey" every two or three years. See Council for the Advancement and Support of Education, "Overall Media Relations Programs—2003 Judge's Report," at http://www.case.org/Content/AwardsScholarships/Display.cfm?CONTENTITEMID=384, accessed.

2. For more on the Indiana University system, see: http://wwwreg.indiana.edu/GradeDistribution/instructor.html, accessed.

3. *The University of Alabama, Student Grade Distribution, 1972–73 through 1975–76*, Tuscaloosa: Office of Institutional Analysis, University of Alabama, 1976), n.p.

4. *American Universities and Colleges* (1973 ed.); *College Blue Book, Tabular Data* (2002 ed.); *Office of Institutional Research, University of Alabama, Student Grade Distribution* (Fall 2000, Spring 2001, Fall 2001, Spring, 2002), unpublished.

5. Office of Institutional Research, Summary, the University of Alabama, Undergraduate Course Grade Distributions, Spring Semesters, 1981 to 1996; *Tuscaloosa News*, September 15, 2002; Valen E. Johnson, *Grade Inflation: A Crisis in College* Education (Ann Arbor, MI: Springer, 2003), 6.

6. Office of Institutional Research, University of Alabama, Student Grade Distribution, Fall 2000, Spring 2001, Fall 2001, Spring, 2002, unpublished.

7. Office of Institutional Research, University of Alabama, Student Grade Distribution, Fall 2000, Spring 2001, Fall 2001, Spring, 2002, unpublished.

8. *The University of Alabama Capstone E-Letter from Interim President J. Barry Mason* 1:4 (October 2, 2002).

9. *Tuscaloosa News*, February 19, 2003; Johnson, *Grade Inflation,* 2–3, 243.

10. *Tuscaloosa News*, September 15, 2002; *Birmingham News* (September 15, 2002).

11. *Tuscaloosa News*, February 19, 2003.

12. Johnson, *Grade Inflation,* 240.

13. Beito to Mike O'Rear, June 30, 2003, O'Rear to Beito, July 1, 2003, Beito to O'Rear, July 1, 2003, William R. Fendley to Beito, July 14, 2003, Beito to Fendley, July 14, 2003, Charles W. Nuckolls to Fendley July 16, 2003, Fendley to Nuckolls, July 16, 2003 (e-mails in possession of author); Charles W. Nuckolls and David T. Beito to Judy Bonner, July 24, 2003; Nuckolls and Beito to Witt, July 26, 2003 (copies in possession of the authors).

Afterword: Focusing on the Big Picture

LESTER H. HUNT

As I have suggested earlier, if you have faithfully read this far, I cannot blame you if you are feeling a little confused. One thing that you can learn from this book, I think, is that some very smart, articulate, and well-informed people hold some sharply contrasting views on issues involving grade inflation and academic standards. Are there *any* truths that we can take as having been established, so that we can go on and debate about the things that remain genuinely debatable? As a step toward answering this question, I wrote up a list that I call (with apologies to the ghost of Karl Marx) "Six Theses About Grade Inflation," and submitted them to the contributors to this volume. Below I list the theses in order, with a few remarks on each. I will conclude with a few more personal observations on where we stand at present, and possible directions for future research and reform.

First, the theses:

1. "Grade inflation" is at best a misleading metaphor. No one disagreed with this, nor did anyone have any qualifications or caveats to offer. In different ways, this point is made by Adelman, who uses the phrase "grade inflation" in ways that adhere very closely to the metaphorical meaning of the phrase, and by Kamber, who favors neutralizing the metaphor by offering an abstract, non-intuitive meaning. Not too surprisingly, Kamber thinks that the term describes a problem that we have today, while Adelman does not. Of course, on this small point of agreement, many theoretical disagreements are balanced, like so many angels on the head of a pin. But that there is this much agreement is surprising—and interesting.

2. Nothing closely analogous to economic inflation can exist in grading. Somewhat to my surprise, Richard Kamber objected to this thesis, as possibly being "dangerously open-ended." As I understand it, his qualm here is based on the fact that there are *some* ways in which grade inflation, when it occurs, *is* like economic inflation. He pointed out that both phenomena can involve a similar sort of causal connection between global changes and resulting local events. Just as economic inflation can mean that the dollars in your wallet, which you worked hard to earn, lose value because of things others are doing far away, so grade inflation can mean that the A that you earned legitimately here at Hometown U. becomes less valuable because of what other people, perhaps at other institutions, are doing. If *they* are giving a lot more As, that can result in people taking your A less seriously. To this Clifford Adelman objected that what Kamber is describing here is devaluation (the currency losing its ability to function as an index of value) rather than inflation: for him, to ascribe inflation to a grade requires a comparison between the grade awarded and the actual value of the student's performance. This illustrates, once again, the conceptual distance between Kamber and Adleman. For Kamber, the judgment that grades are inflated *is* a judgement about the system's functionality—grades lose some of their ability to inform students—and does not require us to say anything about the real quality of the students' performance remaining the same.

For my part, I do not think that Kamber's point indicates a real disagreement with Thesis 2: the fact that grade inflation is in *some* ways like economic inflation does not mean that it is *closely* analogous to it. His point is something more like a warning against being misled by the thesis, rather than a claim that it is not true.

3. *There is at present no received, orthodox view of what the purpose or function of grading is: to inform students, to motivate students, to inform prospective employers and admissions committees, to weed out those who are not learning, or some other purpose.* On this point it was Kohn who was concerned that it might be misleading. That there is literally no orthodox view in the United States of the function of grading may be an inevitable result of the fact that the American educational system is as decentralized as it is. However, he thinks that some functions of grading, as it is actually used, tend to be at least more important than others. In particular, he is concerned about the tendency to use grades

to sort people out into different levels of achievement, and with the use of grades as "extrinsic" motivator.

4. There is often a wide disparity between the grading practices (proportion of As awarded, for instance) from one discipline to another and from one instructor to another. With this no one disagreed, but note their comments on the next one:

5. Such disparities can cause problems. Adelman agreed that disparities between different instructors in the same course or in the same department can cause problems for the accurate assessment of how well the students are mastering the material. However, he doubts that the fact that grades in some courses (such as music performance) are consistently higher than those given in courses in certain other disciplines (such as organic chemistry) cause any problems. He pointed out that we can explain such divergences on the basis of custom and usage. I think the idea here is that if grades in some disciplines are higher because of the local grading conventions, people will take this into account as they interpret the grades. An A in music does not mean the same as an A in organic chemistry. Kohn stated flatly that there is no reason to think that these disparities—apparently meaning any of them—cause problems.

6. The best data do not show a single-direction, nation-wide trend (either upward or downward) throughout the three decades following 1972. Kamber was concerned, as he was with thesis 2, that this statement could be misleading. Though it is true that grades fluctuated during this period—so there is no straight-line direction of change—there is reason to believe that grades were higher at the end of this period than at the beginning, so that the *net* change was upward. He points out that in the newest data with which he is familiar—the National Survey of Student Engagement's 2005 annual report—first-year students reported that 37% of the grades they had received at their current institutions were A or A–.[1] Here Adelman, once again, disagreed. He thought that the NSSE, on this point, cannot be regarded as providing the *best* data, because it relies (as Kamber notes) on information reported by the students themselves. He claimed that when students' self-reported grades are compared with actual transcripts, the former are 0.3 higher than the latter. Further, he thinks that, since the NSSE did not exist before the year 2000, it cannot be compared to any data about the period before that year and, consequently, cannot provide any evidence of

long-term trends. As to Kamber's qualm about thesis 6, perhaps I can save it by introducing a refinement, to the effect that the best data do not show any *continuous* trend in grades in the post-1972 period. Whether this refinement reduces the claim to triviality, I will leave it to the reader to decide.

Perhaps the most impressive thing about the contributors' reactions to my theses is the unanimity about thesis 1. If any one truth emerges from this book, it is that the term, "grade inflation," is a source of confusion. To some extent this is by now obvious. But not all of the confusion is obvious, which is why it persists. "Inflation" is a process, a sort of change. Thus the term, *grade* inflation, and the fact that it is used to point out a problem, suggests that the problem is that grades are higher *than they used to be*. But a careful reading of the essays in this book reveals, I think, that no one here is saying that this is what the problem is. Some of our authors are claiming that grades are *compressed* or *conflated*. Some are alleging and decrying *disparities* in grading. Some are pointing out that grades given are subject to *upward pressures* other than improvement in student work. But no one is saying that change in grading levels is per se a problem. When grade inflation skeptics, like Kohn and Adelman, point out that changes in grading levels are not necessarily bad, or that an extremely literal interpretation of "inflation" does not fit the facts, they are quite right, and their comments to that extent are a positive contribution to the ongoing discussion of these issues. On the other hand, there is a real possibility that these comments, true as they are, can cause confusion, because they suggest that what they are denying is something that their opponents—grade inflation *critics* like Kamber and Biggs—are asserting. As far as I can see, they are not.

Another impressive area of virtual unanimity is on thesis 5, which recognizes the existence of wide disparities between grading practices. It may be that future research and reform in these areas should be more discipline-specific than it has been. Of course, we know what one of the issues will be. Some, like Adelman and Kohn, doubt that disparities between disciplines create problems. Perhaps they are right about that. If students are influenced by grade maximization in their choice of majors, or in choosing between courses in different departments, then intra-disciplinary disparities probably *would* be a bad thing, but it is conceivable that they aren't influenced in that way. Still, there is an issue here, or at least in this immediate neighborhood.

One claim that grade inflation critics have made that I don't think their opponents have really addressed is that, regardless of whether *differences* between grading practices in different disciplines are a problem, there are some disciplines in which the grades given are, in and of themselves, *too high*.

As Mary Biggs points out, grades given by School of Education faculty are often the highest in their college or university, though their students tend to be, but any objective measure, among the weakest.[2] Her comments on this point are corroborated by events at my own institution. In the fall of 1999, the Bruce Beck of the Office of Budget, Planning & Analysis at University of Wisconsin-Madison issued a detailed report, mentioned by Chancellor Wiley in his preface to this book, on grading trends at UW during the nineties.[3] Among his findings (along with a statistically significant upward trend in grades throughout the decade) were dramatic disparities between the grading practices in different disciplines. Figures 1 and 2 show both undergraduate grades and senior grades in the Department of Mathematics and those in the School of Education's important Department of Curriculum and Instruction. Quite aside from the interesting, contested issues here—such as whether the grades in one discipline can be meaningfully compared with those in another, and whether the disparities in themselves can cause problems—one thing seems undeniable: grades cannot be performing the same functions in these two disciplines. In Mathematics, grades are being used to mark the difference between students who have achieved a high level of proficiency in the subject and those who have not, and between those who have mastered the subject to some extent and those who have failed to do so. In Curriculum and Instruction, grades are obviously not being used to perform those functions. Clearly, grades earned in their major subject play little or no role in deciding which of these students will teach the children of Wisconsin in the future and which ones will not. Another fact seems undeniable as well. In Mathematics, grades serve, or to some extent can serve, to mark the difference in levels of proficiency between more advanced and less advanced students: seniors get more high grades and fewer low grades than the average undergraduate student. In Curriculum and Instruction, such differences between levels of proficiency go unmarked and unnoticed by the grading system.

Similar things can be said of at least one other discipline at UW. In October of 2003, Mr. Beck issued a memorandum (occasioned by

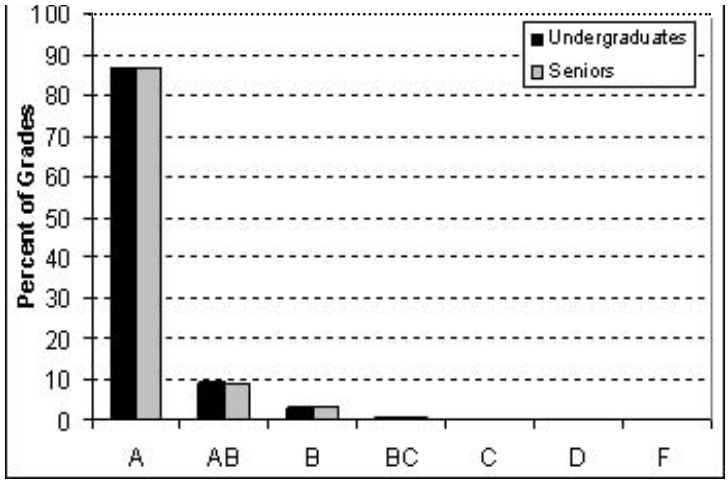


Figure 1. Dept. of Curriculum & Instruction: Undergraduate Grades in Fall 1998.

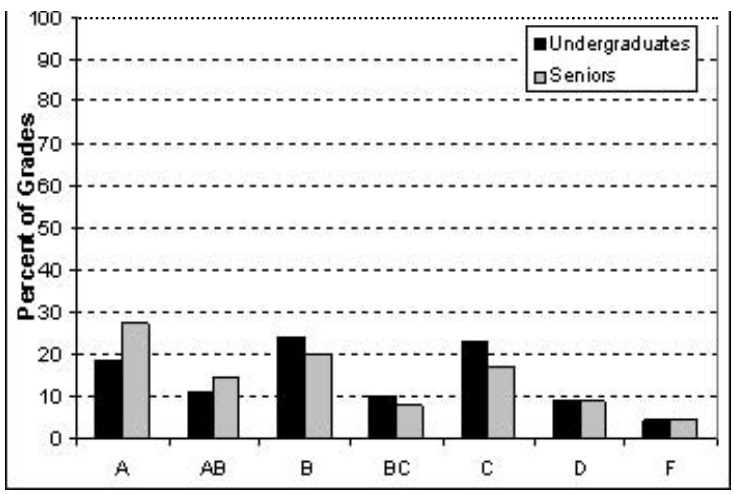


Figure 2. Dept. of Mathematics: Undergraduate Grades in Fall 1998.

the conference from which this book evolved) updating his earlier report on grading at UW.[4] You see part of his results in Figure 3. He compared grades in six different schools and colleges. Further, he only compared grades semester by semester, and only for juniors and seniors. The reason is that the current grades of juniors and seniors have a far stronger tendency to reflect the grading practices of their

Figure 3. Semester GPA by Selected Schools or Colleges (Juniors and Seniors, ACT of 26 and 27 only)

home school or colleges than do the grades of freshmen and sophomores. In addition, in Figure 3 he only compares students with ACT scores of 26 and 27, in an attempt to focus on students with comparable levels of ability. What this graph shows with vivid immediacy is strikingly different rates of change in grades and, more importantly, strikingly different levels in current grade averages. The grades for this group given by the School of Education were one of the two highest. What was surprising, and potentially disturbing, is that Education was surpassed in this respect by the School of Nursing. More important, for present purposes, is the average grades of *all* juniors and seniors

in these disciplines. During the last semester in that period, the upper-class student average for Education was 3.6, a solid A-range grade, while that of Nursing was an even loftier 3.68.[5] Figures 4 and 5 show all the grades given to juniors and seniors that semester in Education and Nursing. As you can see, the overwhelming majority of grades given were As, and nearly all the grades given were As or ABs. I do not know how it is decided which UW Nursing graduates are going to be caring for the sick, but it is difficult to see how grades in their major are going to play a major role in the process of making that decision.

In light of facts like these, it seems that one of the observations one often hears in discussions of grade inflation needs to be drastically revised. This is the claim made by some grade inflation skeptics that today the old C is the new B, the old B is the new AB, and so forth, so that grade inflation, insofar as it exists, has not really changed anything. This is yet another example of reliance on a strong analogy between grade inflation and economic inflation. The idea is that grading changes are macro-phenomena that affect all the micro-economies in law-like ways. What we see may be much more complex and less consistent than that. There are some disciplines in which the new A is something like the old A, and others in which it represents something with no equivalent in the old system. There are also some in which the new C still seems to have some resemblance to the old C, and others in which the old C has simply ceased to exist.

These "new" grades, as I have suggested, cannot serve the same function that grading still seems to serves in mathematics. Is this a good thing or a bad one? On this point there is one truth that seems to me unassailable. It is that we *would* have some reason for confidence that the current situation is not a bad thing *if* it were a result of conscious institutional design, but that we do not have that sort of reassurance. If the people who teach in Curriculum and Instruction, and others that employ the same practices, had consciously decided that grading should not serve the functions that it serves in Mathematics, then we would have the assurance that things are the way they are because people with pertinent knowledge saw reasons why this is how they should be. Further, they might have considered how these practices affect the process of selecting future teachers and nurses, and found reason to think that the process works better, or just as well, when grades in the candidate's major play no role. They might have replaced it with some other mechanism that works at least

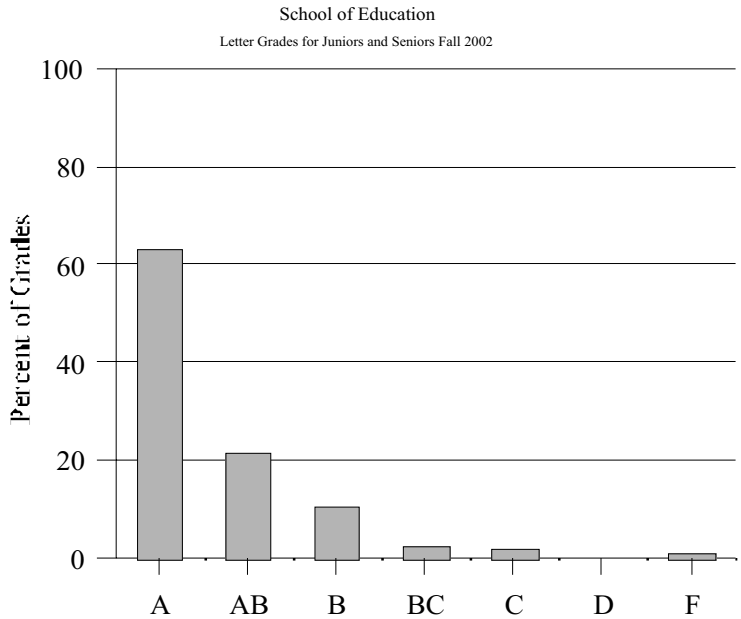


Figure 4. School of Education: Letter Grades for Juniors and Seniors, Fall 2002.

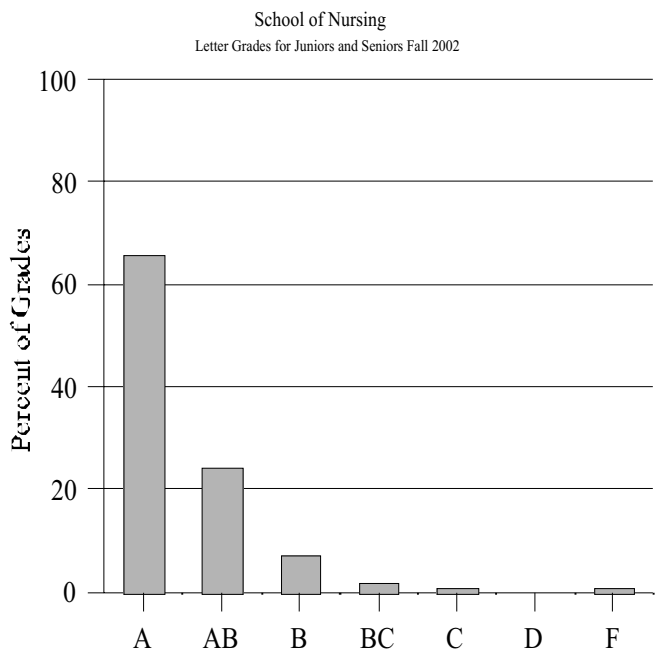


Figure 5. School of Nursing: Letter Grades for Juniors and Seniors, Fall 2002.

as well. Of course none of this has happened. The grading patterns in subjects like this one did not happen because there was a discussion of whether they should happen or not: they just happened.

This of course raises the interesting question of why such things do happen, and whether the reasons why they did happen are of the sort that can reassure us that the system might well be working as it should. On this point I would like to abuse the privileged position of editor of this volume one more time, and offer a possible explanation. It is admittedly pure speculation, but it also has I think a fairly strong facial plausibility. I suppose I should admit, though, that my explanation is not of the reassuring sort. In fact, it implies that the practice of grading, as it exists today in the United States, constitutes an institutional arrangement that may be fundamentally flawed. The flaw is that the system requires one person, the instructor, to carry out two radically different sorts of activities, ones that rest on contrasting states of mind: namely, teaching and grading.

To some people, this claim, that there is to some extent a conflict between teaching and grading, will sound very foolish. After all, teaching requires that we give students feedback on how they are doing. If I were teaching you to play the violin, I would have occasion to say things like: "No, your left wrist should be straight . . . the French bow-grip shouldn't be all clenched up like that . . . in the *piano* passages your bow is way too close to the bridge, move the bow over the fingerboard." But is this sort of feedback *grading*? I don't think so. Grading has features which these sorts of comments do not have, whether on behalf of the University of Wisconsin we are grading a student's performance on the violin as A, B, C, D or F, or whether on behalf of the Department of Agriculture we are grading a pineapple as U. S. Fancy, U. S. No. 1, U. S. No. 2, or U. S. No. 3. One such feature is that grading consists in assigning the thing being evaluated to one of a pre-determined set of sharp-edged categories: although the criteria for consigning a performance to the A-category, or a fruit to the Fancy category, may be fuzzy, there is nothing fuzzy about the resulting categorial status. A performance that has been graded "A" clearly has not been graded "B." Second, the categories used in grading are hierarchical: each one is clearly ranked either above or below each of the others. To grade is to police the boundaries of these categories, it is to guard each category from being invaded by things that belong in lower categories. The idea that a piece of fruit that is unfit for human consumption

might slip through and be graded Fancy—that is a fruit grader's worst nightmare.

Grading is tough work, but fortunately there are people who are comfortable doing it. I suspect that they are the sort of people, perhaps with a different set of skills, who would enjoy being judges or border guards. Indeed, I have said in effect that a grader *is* a sort of border guard. But I don't think that graders would like to be teachers, nor that teachers would like to be graders. Teaching is one of "the helping professions." We become teachers because we want to help people to grow, and not because we want to put them in their place—which is the point of any sort of grading. A teacher does not police limits, rather they help you to violate them. Try to imagine Socrates grading Phaedrus. Isn't the very idea absurd? Why?

I think this is a deep question, and I don't pretend to have a complete answer to it. Suffice it to say that a true teacher will always to some extent detest the necessity of grading, while true graders (people who are born to guard boundaries) will to some extent despise teaching, the helping spirit behind it, and the need for encouragement and help to which it ministers. Of course, the current educational system requires grading, and thus requires people to be both of these things at the same time and in relation to the same people. One perhaps inevitable result will be that people will use the institution of grading for teaching-purposes and not for grading-purposes. That is, they will use it to encourage and support their students. However, this nurturing tendency is a personality-trait that some individuals (including some individual teachers) might have in much greater abundance than others do. Thus we sometimes have people like Mary Biggs' colleague, who has two grades: "A, good job" and "A, nice try," in the same department with people who award grades with the intention of sorting students into levels of achievement. Further, different academic disciplines appeal to different sorts of people. If we just suppose that this nurturing psychological profile is one that can characterize the sorts of people who are attracted to one discipline far more that it characterizes those attracted to another, we can explain why entire disciplines award grades in the student-nurturing way, as apparently do the UW nursing and education faculties. Since the explanation I am offering includes a distinctive analysis of what grading is—that it involves policing the boundaries between achievement-categories—has a certain implication that might easily escape notice. It implies that people who do such things are not

applying a special philosophy or personal approach to grading. They are simply avoiding grading because they are averse to it. What they are doing is not grading at all.

More obvious, I suppose, is the fact that the explanation that I have given for this fact, that grading is a practice that is fading away in certain quarters, is not the sort that reassures us that the phenomenon is not a bad thing. If my explanation is right, it is happening, not because some people have reason to think that it is a bad thing, but because they find it distasteful. The fact that they find it distasteful may perhaps mean that it clashes with values, feelings, or perceptions that are an indispensable component of their professional activity, but this would only show, at most, that grading should not be done *by them*, not that it should not be done (perhaps by someone else).

At this point I am sure many people would want to deny the psychological dichotomy I have sketched between teaching and grading. One might well agree with me, that what people are doing when they give enormous numbers of As in order the make students feel good about themselves is not grading, but go on to point out that this activity is not properly called teaching, either. Of course there is a certain conflict, one might say, between teaching and grading, but one is not a *good* teacher unless one handles the conflict appropriately. After all, a good teacher conveys to students a sense of the limits of what they have attained so far, and vision of the greater achievements that still lie ahead of them. Without this, teaching becomes a process without a goal or point. At this point, a moralist like Mary Biggs might add that, though there is a sort of psychological tension between the growth-nurturing and boundary-policing functions of the great teacher, one of the functions of sound moral character is to handle conflicts like this in a properly balanced way, without caving in to the temptations of one side or the other. The function of virtues like courage, temperance, self-control, and fortitude is to enable us to handle such conflicts as this one. In the world of teaching and grading, there is no substitute for strength of character.

I do not deny any of this—except for the very last statement. To some extent, there *are* substitutes for good character, and it is wise to make some use of them. Our own system of government, with its system of checks and balances, is designed to work (at least, not collapse into dictatorship) even if the people who run it are not virtuous. Surely, a

system that only works if it is staffed by people who are of sound character is doomed from the beginning. Perhaps it is time to ask whether the present institutional arrangement requires *too much* strength of character from the people who actually operate it.

Surely there is one thing that cannot be denied: the system does require more virtue than it did a generation ago. One of the things that makes honest grading difficult for a teacher is the fact that teaching at its most intense is a very personal relation between individuals, while grading is an equally impersonal relation. If I am your violin teacher, we have spent the term making music together. That is very personal, or can easily become so. This can make it very difficult to turn at the end of the term and officially brand your musical achievement as mediocre, even if that does represent my honest opinion. When I was an undergraduate, in the middle sixties, this difficulty was eased somewhat by the relative formality—in other words, *im*personality—of the relations that then prevailed between faculty and students. Students addressed faculty as Professor or Doctor, and faculty addressed students as Mr. or Miss. This impersonality weakened the growth-nurturing aspect of teaching, but it strengthened the position of the grader. In this way, the institutional arrangements leaned against the direction in which weak human nature is prone to err, taking some of the burden off of individual strength of character. Of course, this particular arrangement was in effect abolished in the late sixties.

Then, too, the position of the grader was strengthened by the fact that the relationship between teacher and student was one of utterly asymmetrical power. The faculty member was both teacher and evaluator of the student—two functions that represent relative power—while the student was in comparison passive and powerless. The teacher graded the student, and the student did not grade the teacher. This arrangement was also more or less demolished in the late sixties, by the innovation that I discussed in my contribution to this volume ("Grading Teachers: Academic Standards and Student Evaluations"): reliance on student evaluation of teaching. As a matter of fact, this innovation did much to *reverse* the old asymmetry of power. Unlike the grades that faculty give to students, the student rating of professors is routinely done anonymously, so for students, unlike faculty, there is no cost for issuing low grades. Further, while grades seem to have less and less impact on students' futures, compared to test scores and

other factors, student ratings of teachers have a routine, annual effect on the faculty member's income, for these ratings are almost the only source of information about teaching effectiveness that is relied upon by the people who decide merit pay-increases for faculty.

In this new environment, grading represents a constant and genuinely onerous test of the teacher's courage and integrity. Are there any realistically imaginable reforms that might lighten this burden? I doubt that reversing the egalitarian reforms of the sixties is one of them. No doubt, faculty shed their former position of pompous dignity and authority because they did not want it, and they probably would not want to take it up again, even if there were some conceivable way to do so. In all probability, there is no turning back the clock by revoking the egalitarian reforms of the sixties. One possible reform, one that arguably would not be counter-revolutionary, would be one that I suggested earlier: namely, we might introduce serious, sustained, *faculty* evaluation of teaching. Students could go on filling out their evaluation forms, and faculty could consult them for useful feedback, but if teaching is evaluated by experts, there will be less reliance on student evaluations for merit-pay-raise purposes and consequently less pressure to please (and, worse yet, avoid offending) students. Another, more radical, possibility would be to consider institutionally separating the process of teaching from that of evaluating students. Teachers could continue to give students feedback of the improving-your-bow-grip variety, while the task of sorting them out into levels of achievement can be done by means of department-wide exams or other assignments that are evaluated by external examiners.

As higher education has evolved, in recent generations of practitioners, both the teaching and the evaluation of students have changed significantly. Perhaps the current generation, and the ones that are coming along, should consider taking control of these changes, and perhaps they should consider changing the way teaching and evaluation are related to each other.

NOTES

1. This report, he informs me, is available on-line at http://nsse.iub.edu/nsse_2005/index.cfm.

2. See the works cited in footnote 6 in her "Fissures in the Foundation: How Grade Conflation Could Happen," this volume.

3. The entire report can be found at http://www.apa.wisc.edu/grades/UG_grades.htm.

4. Memorandum dates October 9, 2003 and signed by Senior Policy and Planning Analyst Bruce Beck.

5. This information was supplied to me by Bruce Beck on August 7, 2006.

This page intentionally left blank.

Contributors

Clifford Adelman (Senior Research Analyst, U.S. Department of Education) has conducted a number of longitudinal studies for the USDE, including "Answers in the Tool Box: Academic Intensity, Attendance Patterns, and Bachelor's Degree Attainment," and has also followed statistical trends in grade inflation.

David T. Beito (Associate Professor of History, University of Alabama, Tuscaloosa) is the author of two books and many articles about American history. He has also written op-ed pieces on grade inflation and, with Charles W. Nuckolls, a study of grade distortion at the University of Alabama.

Mary Biggs (Professor of English, The College of New Jersey) has taught at both the undergraduate and graduate levels and in schools of librarianship, and currently teaches in an English department. She has published on a wide range of subjects including higher education, intellectual freedom, women studies and literature.

Harry Brighouse (Professor of Philosophy, University of Wisconsin, Madison) is the author of two recent major reports on education for left-of-center think-tanks as well as *School Choice and Social Justice* (2000). He has also commented widely in the British national press on issues in education.

Lester H. Hunt (Professor of Philosophy, University of Wisconsin, Madison) has written extensively on ethics and political philosophy.

He is the author of *Niezsche and the Origins of Virtue* (1990) and *Character and Culture* (1998).

Richard Kamber (Professor of Philosophy and Chair of the Department of Philosophy and Religion at The College of New Jersey) served for sixteen years as a college dean and vice president. His publications include articles on higher education, film, and the Holocaust, as well as books and articles on a variety of philosophical topics.

Alfie Kohn (Author and Scholar, Belmont, Massachusetts) is the author of eight books and many articles on education and human behavior including, most recently, *The Case Against Standardized Testing* (2000). He was recently described by *Time* magazine as "perhaps the country's most outspoken critic of education's fixation on grades [and] test scores."

Charles W. Nuckolls (Professor of Anthropology, University of Alabama, Tuscaloosa) is a psychological and medical anthropologist. His primary interest is the theory of culture, and in integrating cognitive and psychoanalytic approaches to everyday thinking.

Francis K. Schrag (Professor of Educational Policy Studies and Philosophy, University of Wisconsin, Madison) has published many articles and two books on diverse topics in education and social philosophy, including *Back to Basics: Fundamental Educational Questions Reexamined* (1995).

Index

This page intentionally left blank.

EDUCATION

Grade Inflation

Academic Standards in Higher Education

Lester H. Hunt, editor

"As state and federal agencies begin to talk about accountability for universities, the topic of grade inflation could become even more politicized. This timely book addresses a topic of significant public interest and does it well. The fact that the contributors disagree, take different approaches, and address different aspects of grade inflation is a virtue."

— Kenneth A. Strike, author of *Ethical Leadership in Schools: Creating Community in an Environment of Accountability*

"This book encourages academic communities to engage in constructive debate over their professional responsibilities as evaluators of student academic work. Its greatest strength is that it presents disparate perspectives on the complex topics of grading and grade inflation. The contributors are in a real sense engaged in a discussion on the subject, which makes the book refreshing and intellectually stimulating."

— Matthew Hartley, University of Pennsylvania

ISBN: 978-0-7914-7497-6

EAN 9 780791 474976 90000

State University of New York Press
www.sunypress.edu