

What Are Grades Made Of?

Alexandra C. Achen and Paul N. Courant

The term “grade inflation” covers a multitude of phenomena, some of which are even alleged to be sins. Continuing increases in average grades have been widely documented in many universities over the last several decades (for example, Sabot and Wakeman-Linn, 1991; Johnson, 2003). Conversely, cases of grade deflation are rare and short-lived, although in some settings, such as first-year law courses, some universities have held to a strict curve. Also widely documented, and often associated with grade inflation, are systematic differences in grade levels by field of study, with a common belief that the sciences and math grade harder than the social sciences, which in turn grade harder than the humanities—and that economics behaves more like the natural sciences than like the social sciences. The general persistence of these relative differences in grades seem to us to be more interesting and more difficult to explain than the persistence of modest grade inflation in general, and they are the principal focus of this paper. Why, for example, should average grades in English be much higher than average grades in chemistry? And what is going on when a department’s grading practices change markedly relative to other departments?

We begin with an overview of some evidence on grade inflation by department and course level, focusing in particular on detailed data that we have from the University of Michigan. Grades in undergraduate arts and sciences courses at the University of Michigan have, with a few exceptions, been rising slowly and steadily since at least 1992. But our main focus is to explore some possible reasons for the highly (but not perfectly) stable differences in the grading practices of depart-

■ *Alexandra C. Achen is a Ph.D. student in Sociology and the Gerald R. Ford School of Public Policy, University of Michigan, Ann Arbor, Michigan. Paul N. Courant is Arthur F. Thurnau Professor of Economics and Harold T. Shapiro Collegiate Professor of Public Policy at the University of Michigan, Ann Arbor, Michigan. Their e-mail addresses are <achena@umich.edu> and <pnc@umich.edu>.*

ments. Perhaps surprisingly, we uncover a story that is much richer and more interesting than some variant of “the sciences (which are virtuous or mean, depending on your point of view) grade tough, and the humanities (which are the opposite, again depending) grade easy.”

Our basic story is fairly simple. Grades are an element of an intra-university economy that determines, among other things, enrollments and the sizes of departments. Departments supply courses and students demand them, although the payment from students to faculty is mediated by the university administration, and there are also nonpecuniary rewards and costs associated with teaching. Departments generally would prefer small classes populated by excellent and highly motivated students. The dean, meanwhile, would like to see departments supply some target quantity of credit hours—the more the better, other things equal—and will penalize departments that don’t do enough teaching. In this framework, grades are one mechanism that departments can use to influence the number of students who will take a given class. But both the costs and consequences of different grading policies vary systematically across departments and courses. Grading is always at least somewhat costly, but the cost is greater the greater are the opportunities for students to quarrel with the fairness of the grading standards and methods that faculty use. On the demand side, some courses have close substitutes while others do not, and one would expect the grade-elasticity of demand to behave in the usual way.

This framework leads to several hypotheses about relative grades across departments and courses. First, the distribution of grades is likely to be lower where courses are required, and where there are agreed-upon and readily assessed criteria—right or wrong answers—for grading. By contrast, departments that evaluate student performance using interpretative methods will tend to have higher grades, because using these methods increases the personal cost to instructors of assigning and defending low grades. Second, upper-division classes are likely to have higher grades than lower-division classes, both because students have selected into the upper-division courses where their performance is likely to be stronger and because faculty want to support (and may even like) their student majors. Third, grades can be used in conjunction with other tools to attract students to departments that have low enrollments and to deter students from courses of study that are congested.

We find some evidence in support of each of these patterns. As it happens, the consequence of the preceding tendencies is that, indeed, the sciences (mostly) grade harder than the humanities. But there are some surprises. For example, consistent with our framework but not consistent with the notion that the humanities grade softer than the sciences for some intrinsic reason, we find that at Michigan introductory physics and chemistry labs grade much easier than second-year French courses.

We find relative grades to be both more interesting and more amenable to analysis than the low rate of general grade inflation. Inflation, we expect, arises from two complementary features of the landscape: for any instructor in any course, grading a little more softly than expected is costless and it makes students happy. The instructor may gain some benefit in teaching evaluations (Johnson,

2003), but even if not, the opportunity to (in effect) print money is one that at least some instructors will find appealing. As long as some faculty respond to this opportunity, others in the department will be under pressure to adjust to the new norms for grades, and at least some other departments will endeavor to follow this trend in order to maintain market share and perhaps also to avoid the unpleasantness of widespread student grumbling. This story is hard to verify or to refute, in part because general grade inflation has proceeded without interruption for so long, but the key ingredients are surely in place.

We conclude with a discussion of implications for further research and for academic policy. We argue that differential grading standards have potentially serious negative consequences for the ideal of liberal education. At the same time, we conclude that any discussion of a policy response to grade inflation must begin by recognizing that American colleges and universities are now in at least the fifth decade of well-documented grade inflation and differences in grading norms by field. Current grading behavior must and will be interpreted in the context of current norms and expectations about grades, not according to some dimly imagined (anyone who actually remembers it is retired) age of uniform standards across departments. Proposals that attempt to alter grading behavior will face the costs of acting against prevailing customs and expectations, whether in altering pre-existing patterns of grades across departments within a college or university or in attempting to alter grades in one institution while recognizing that other universities may not change.

Grade Inflation and Relative Grades: A Look at the Data

Data on grades are sparse. At many colleges and universities, such data are not publicly available—at least not in any form that might risk leading to controversial findings—and no systematic data exist for comparing course grades across colleges and universities. This study is therefore based on current data from a single institution—the University of Michigan—and it could not have been undertaken without the generosity of the University’s registrar. Universities are different from each other, of course, and we do not expect that Michigan (or anywhere else) is typical in its details. Indeed, part of our story is that some differences in grading practices are due to differences in norms and relevant structures (like course distribution requirements) and these will surely vary across universities. At the same time, broad patterns that have been documented elsewhere hold generally at Michigan, giving us some confidence that our findings have some generality.

The Michigan data that we employ are very rich, covering 25 departments from Fall 1992 through Winter 2008. (We leave out a number of small departments and departments that change identity over the period in confusing ways.) We limit our study to Michigan’s College of Literature, Science, and the Arts (LSA), which is the undergraduate liberal arts college, with about 16,000 students of its own and which provides lower-level instruction to many thousands more in other undergraduate programs. Notably, LSA provides the basic science and math courses to students in

the College of Engineering, which is the second-largest undergraduate College at Michigan. These Michigan data allow us to distinguish between upper-division courses, which are mostly taken by majors and others who have shown interest or aptitude for a field of study, and lower-division courses, most of which are introductory. We are also able to identify courses that are required in order to complete desirable or popular courses of study. No courses are required for all students, but certain courses can't be avoided if students want to meet pre-med requirements, or graduate from the engineering college, or gain entry to the business school, or pursue a number of other desirable programs.

The data show continuing grade inflation throughout the period 1992–2008. Weighted by credit hours, average annual rates of grade inflation (measured absolutely on the standard four-point scale) at the University of Michigan over the period for which we have data are 0.011 for lower-division (freshman/sophomore-level) courses and 0.009 for junior/senior courses. At this rate, a B (grade of 3.0) would rise to a B+ (grade of 3.3) over about 30 years. All but two departments—statistics and communication studies—exhibit grade inflation in their lower-division courses. Four departments have declining grades in their upper-division courses: biology, comparative literature, statistics, and physics. Of these, only biology and statistics are meaningfully different from zero. The most rapid inflation rate in the group is in lower-division geology courses, at 0.026 of a grade point per year. The second most rapid is upper-division art history at 0.025 of a grade point per year. If these departments have anything else in common, it does not leap to mind. There is very little general to be said about grade inflation over the period—except that it appears almost everywhere.

The inflation rates from 1992 to 2008 in our Michigan data are somewhat lower than those documented by Sabot and Wakeman-Linn (1991) for Williams College and seven other institutions over the period 1962–63 to 1985–86. Given the fact that 4.0 is an upper bound on the grade point average and that a number of departments are above B+ as of 2008, the slowing down of inflation may in part be mechanical. While it is plain that departments differ in their grading practices, even departments that give few C's and below continue to produce both A's and B's, presumably to provide meaningful rewards and signals. More interesting than continuing inflation, however, is the stability of departments relative to each other. The first two columns of Table 1 show average grades in introductory courses across seven departments for an average of seven institutions in 1985–86, while the second column shows grades for intro courses at Williams in 1985–86. The third column shows our University of Michigan data for the same departments over the 1992–1994 period, and the final column shows Michigan data for the 2005–2007 period. With the exception of psychology at Williams, which was relatively low-grading in the mid-1980s, the departmental ranks over time never change by more than one position. Math grades the hardest, although even in math, grades in lower-division courses are rising over the period. Economics comes in second or third in all of the columns. English grades high and keeps grading higher. Also striking is continuance of grade inflation. In all cases in Table 1, the Michigan grades in the early 1990s are above the averages reported by Sabot and Wakeman-Linn for the mid-

Table 1
Grades in Seven Departments, Selected Periods and Places
(ranks in parentheses)

<i>Department</i>	<i>Average from seven institutions, intro courses (1985–86)</i>	<i>Williams College, intro courses (1985–86)</i>	<i>University of Michigan, lower-division courses (1992–94)</i>	<i>University of Michigan, lower-division courses (2005–2007)</i>
Chemistry	2.66 (2)		2.95 (3)	3.03 (3)
Economics	2.81 (3)	2.67 (2)	2.83 (2)	2.92 (2)
English	3.12 (7)	3.13 (6)	3.19 (7)	3.39 (7)
Mathematics	2.53 (1)	2.61 (1)	2.83 (1)	2.87 (1)
Philosophy	2.99 (5)	2.94 (4)	3.15 (6)	3.30 (5)
Political science	2.95 (4)	3.10 (5)	3.11 (4)	3.19 (4)
Psychology	3.02 (6)	2.71 (3)	3.13 (5)	3.36 (6)

Note: The data for 1985–86 (first two columns) is reproduced from Sabot and Wakeman-Linn (1991).

1980s, and in all cases the Michigan grades from our data in 2005–2007 are higher than at Michigan in the earlier period.

Table 2 also shows general stability in lower-division relative grades across departments at Michigan, but when we look at 25 departments rather than the seven shown in Table 1, we see more variation. Two of the most interesting cases, which we discuss at greater length below, are geology, which goes from having very low average grades (relative to a median for the period of 3.13) to about average in the second period (median 3.30), and communications, which moves from above average to well below. The most extreme behavior is in statistics, which goes from the middle of the pack to being one of the lowest-grading departments.

As can be seen in Table 3, in the most recent period, most departments behave quite differently in upper-division courses than in their lower-division offerings. With the exception of linguistics, American culture, geology, sociology, and philosophy, none of which is a low-grading department, grades in upper-division courses are higher than those in lower-division courses in all departments. The median difference over all 25 departments is 0.12, and it tends to be larger the lower are lower-division grades (the simple correlation is -0.54). Math is not *the* hardest-grading department in the upper division, although it is one of the hardest. Many observers may find it surprising that philosophy, chemistry, physics, Afroamerican and African studies, and sociology are within 0.06 of a grade point of each other in the upper division—with philosophy the lowest-grading of the group. While a good deal of dispersion remains, departments behave more like each other in upper-division courses than they do in lower-division courses. The range in average grades between the highest- and lowest-grading department is 0.64 for upper division and 0.73 for lower-division courses. Dropping the bottom three departments, the range is 0.49 for upper and 0.60 for lower.

Finally, some departments show differences in grade inflation rates between upper- and lower-division courses. Inflation is modest in lower-division math courses, but much higher in the upper division. The converse is true for physics. Chemistry has high inflation at the upper level, but not in the lower division. Two

Table 2
Lower-Division Grades at the University of Michigan, 1992–2007
(ordered by average grade in 2005–2007)

<i>Departments</i>	<i>1992–1994</i>	<i>2005–2007</i>	<i>Inflation (rank)</i>
Mathematics	2.83	2.87	0.05 (4)
Statistics	3.14	2.89	−0.25 (1)
Economics	2.83	2.92	0.09 (8)
Physics	2.84	3.00	0.16 (15)
Biology	2.92	3.02	0.10 (10)
Chemistry	2.95	3.03	0.08 (6)
French	3.04	3.06	0.02 (3)
Astronomy	2.95	3.13	0.18 (17)
Communications	3.28	3.13	−0.14 (2)
Political Science	3.11	3.19	0.08 (5)
Spanish	3.08	3.23	0.15 (13)
Geology	2.94	3.29	0.34 (25)
Philosophy	3.15	3.30	0.14 (12)
Afroamerican and African studies	3.09	3.30	0.21 (19)
History	3.14	3.31	0.16 (16)
Sociology	3.21	3.35	0.14 (11)
Psychology	3.13	3.36	0.23 (22)
English	3.19	3.39	0.20 (18)
Anthropology	3.09	3.40	0.31 (24)
History of art	3.19	3.41	0.22 (21)
Women’s studies	3.34	3.44	0.09 (9)
Linguistics	3.37	3.53	0.16 (14)
German	3.26	3.56	0.30 (23)
American culture	3.36	3.57	0.21 (20)
Comparative literature	3.51	3.60	0.09 (7)
Median grade for the period	3.13	3.30	

departments, biology and statistics, show significant deflation over the period in their upper-division grades.

Why Grades Vary Across Departments and Classes

We take it as given that faculty give grades in part to provide rewards, punishments, and signals of academic performance, and that faculty tastes regarding the importance of giving grades vary. But plainly other motivations are at work as well, or else we would not observe ongoing grade inflation and the occasional changes in relative grading practices that have been documented here and elsewhere.

One can imagine a department determining an optimal size and distribution between lower-division and upper-division courses, in response to some combination of internal politics and external pressures from university administrators. The department would then adjust its grading policy and other relevant policies,

Table 3

Upper-Division Grades at the University of Michigan, 1992–2007*(ordered by average grade in 2005–2007)*

	1992–1994	2005–2007	Inflation (rank)
Biology	3.21	3.03	–0.18 (1)
Statistics	3.18	3.04	–0.15 (2)
Economics	3.03	3.06	0.02 (7)
Mathematics	2.99	3.18	0.19 (20)
Geology	3.09	3.26	0.17 (16)
Philosophy	3.18	3.27	0.09 (11)
Chemistry	3.10	3.28	0.18 (17)
Physics	3.26	3.28	0.02 (6)
Afroamerican and African studies	3.25	3.30	0.06 (8)
Sociology	3.32	3.33	0.00 (5)
History	3.24	3.37	0.14 (14)
Communications	3.35	3.40	0.06 (9)
Political science	3.20	3.40	0.20 (21)
English	3.25	3.43	0.18 (19)
Astronomy	3.46	3.44	–0.02 (3)
Linguistics	3.41	3.47	0.06 (10)
Anthropology	3.32	3.50	0.18 (18)
History of art	3.22	3.53	0.31 (24)
American culture	3.38	3.53	0.16 (15)
Psychology	3.30	3.55	0.25 (23)
Women’s studies	3.49	3.58	0.10 (12)
German	3.49	3.61	0.12 (13)
French	3.40	3.63	0.23 (22)
Spanish	3.31	3.63	0.32 (25)
Comparative literature	3.67	3.67	–0.01 (4)
Median grade for the period	3.24	3.40	

including requirements for the major and for particular courses, in an effort to attain that size and distribution of courses. There is powerful evidence that students (especially weaker students) tend to be attracted to courses that have high average grades (for example, see Sabot and Wakemann-Linn, 1991, Johnson, 2003; and the article in this issue by Bar, Kadiyali, and Zussman). Thus, grades can be used as policy instruments to influence enrollments. The effects of grades on enrollments, however, will vary depending on other features that affect the demand for enrollment in a department’s courses. In particular, we would expect the consequences of grading policies to be quite different in required courses than in electives.

Required and Introductory Courses

From the student point of view, introductory courses come in three broad flavors: 1) There are courses that students take to learn something about a field that they do not know well, perhaps with an openness to taking more advanced courses if things go well, perhaps not. 2) There are courses that students choose to fill distribution requirements, though they typically have a good deal of choice about

how to fulfill the requirement. 3) There are courses that are required in order to fulfill university requirements or prerequisites for majors and courses of study. This last group is large: Engineering students are required to take calculus and physics; almost everyone is required to take English composition; business majors are required to take introductory economics; and psychology majors and others are required to take introductory statistics. Pre-meds must take math, physics, biology, and chemistry. It is worth noting that when particular courses are required to follow a later academic path, many—often most—of the students in those courses will be there to fulfill a requirement that is not imposed by the department that teaches the course.

Each of these types of introductory course fits differently into the economy of the university. Departments that teach a lot of students in courses of the third type (required for popular courses of study) may choose to grade easy or hard in these courses, but they can be assured of a large core of grade-inelastic enrollments in either case. To be sure, some pre-med hopefuls will change their minds in response to poor performance in introductory chemistry or biology or physics, but as long as a fairly large number of students come to the university with strong and specific professional ambitions, the courses required to fulfill those ambitions will have high and inelastic demands. In contrast, courses that may be used to fill distribution requirements but that are not themselves specifically required for a course of study are likely to face more grade-elastic enrollment demand. Thus, depending on other circumstances in the department, grades in such courses could be used as one instrument to affect enrollments. To the extent that these courses also serve as gateways to further work in the specific field, grading standards may have a multiplier effect, positive or negative, on demand for a department's more advanced offerings.

Strikingly, all of the low-grading departments at Michigan have large required courses. Math, physics, economics, biology, and chemistry are consistently among the lowest-grading departments, and each has courses that are required for courses of study necessary for the professional ambitions of large numbers of students.¹ One might argue that the common attribute of introductory courses in these fields is that they involve quantitative work, but so do geology and astronomy, which grade much lighter in their lower-division courses, and which compete to attract enrollments from students who are merely trying to meet a distribution requirement in science. The anomaly here is English 125, required freshman composition, where the average grade in the recent period is B+, no lower than other English courses notwithstanding the fact that the grade elasticity of demand for English 125 is surely close to zero—(almost) no one takes freshman composition for the love of it.

Table 4 shows both that required courses account for more than half of lower-division enrollments in all departments that have required courses except for

¹ Interestingly, this logic applies to economics twice because introductory microeconomics is required for the undergraduate business school degree and because economics is one of the few majors in the College of Literature, Science, and Arts whose major-specific training is relevant to obtaining good jobs with a BA degree.

Table 4
**Share of Credit Hours in Required Courses, by
 Department, University of Michigan, 2005–2007**

<i>Department</i>	<i>Lower-division credit hours</i>	<i>Share in required courses</i>
Mathematics	31,742	0.765
Physics	19,737	0.912
Chemistry	29,318	0.878
Economics	13,389	0.651
English	26,492	0.580
Biology	19,460	0.399

Note: Statistics is omitted from the table because the required courses are in the upper division.

biology, and also that the absolute numbers of credit hours in these courses are large. If anything, the numbers understate the importance of required courses in at least some departments. Students intending to apply to medical school may strengthen their applications by taking several courses in biology, but we count only the first semester of biology as “required.” In math, most of the lower-division credit hours not counted as “required” are taught in honors sections that parallel the coverage of the required courses. In economics, we code introductory micro as “required,” because it is necessary for admission to the undergraduate business school degree as well as to the economics major. Almost all of the remaining lower-division credit hours are in introductory macro, which is required for the economics major (which is attractive to vocationally-oriented students in itself) but not for admission to the business school.

To explore further the sources of low grades in required introductory courses, we looked at a set of language courses that act like required courses in that they are essential to completing valuable courses of study. Graduation from the College of Literature, Science, and the Arts (LSA) with a BA or a BS requires successful completion of four semesters of a foreign language.² Thus, the fourth semester in any language should act much like a required course. Students are often able to test out of some or all of the required language study based on work completed in high school. It is common for students to arrive at the university with the equivalent of two or three semesters of a foreign language completed, and for these students, the fourth-semester course in that language is by far the easiest path to meeting the language requirement. Thus we would expect languages that are commonly taught in U.S. high schools—Spanish and French—to have relatively grade-inelastic demand for the fourth-semester course in LSA. We examine this in Table 5.

In French and Spanish, the grades in the final “required” course (232) are in the same territory as introductory science grades (relatively low). By contrast,

² LSA also gives another bachelor’s degree, the Bachelor of General Studies (BGS), that does not require a foreign language. Fewer than 5 percent of LSA graduates receive the BGS, implying that more than 95 percent complete four semesters of a foreign language.

Table 5
Grades in Selected Language Courses at the
University of Michigan, 2005–2007
(ordered by average grade in the fourth semester)

	<i>4th Semester</i>	<i>Upper-level</i>
French	2.91	3.63
Spanish	3.16	3.63
Japanese	3.35	3.62
Italian	3.46	3.32
Russian	3.51	3.61
Chinese	3.58	3.62

Note: Upper-level language courses in both Chinese and Japanese are listed through the Asian languages department, so we report a single average for grades in upper-level courses in this department.

Italian, something people don't take in high school, but which is also a Romance language, taught in the same department as French and Spanish, looks more like English 125 (grades are relatively high). Thus, within Romance Languages, it would seem that French and Spanish departments see no reason to attract students who are already on the path to meet the language requirement via those languages, while Italian, which is not available at most U.S. high schools, treats its lower-division students more like majors or other upper-division students. It is at least plausible that this behavior is part of a strategy to attract students who are starting fresh in their language sequence and have a large number of potential languages from which to choose, both in terms of introductory courses and majors.

Consistent with this interpretation, average grades for Chinese, Japanese, and Russian (languages most students can't take in high school) look more like those for Italian than for French or Spanish. What we don't know, of course, is the extent to which students are committed to these languages for reasons that are independent of grades. It may be that students in the less common languages have a salutary combination of interest and talent, and hence tend to get good grades for that reason.

Upper-Division and Lower-Division Courses

All of the departments that grade hard at the introductory level have many students enrolled in often-required courses, or courses that face persistent congestion and excess demand, or both. At the same time, we find that most departments (and almost all of the departments that have low average grades in introductory courses) treat students much better in the upper-division courses.³ For both upper

³ There are two interesting exceptions in the Michigan data. In economics, it turns out that the grade distribution in intermediate microeconomics and macroeconomics, which are upper-division courses, are even lower than that in the introductory courses. If we treat the intermediate theory courses as lower-division, upper-division economics would move to 3.15 from its current level of 3.06, and the

and lower divisions, we see systematic and often persistent differences in grades by field, as well as continuing increases in grades in almost all fields at all levels. We also see a general pattern in which upper-division grades are higher than those in the lower division.

This pattern makes some intuitive sense. By and large, students in the upper-division courses signal both interest and some talent merely by being there. In advanced courses in math and science, students cannot get to the upper division without having shown some aptitude for the field. Students know something about what they like and are good at—having learned in part from the grades they received, relative to their peers, in the introductory courses. Generally then, compared to the case with introductory courses, the students in upper-division courses tend to be more interested and competent in the subject matter. Even if students in upper-division courses were graded on the same standards as in the introductory courses, on average they would do better because they are a biased (in a good way) sample of the population that began in the introductory courses in the same fields.

Moreover, from the perspective of the faculty, precisely because upper-division students are likely to be interested in the field, upper-division courses are more fun to teach, and the dynamics of the class are likely to be more salutary than in the lower division. This should lead to a secondary positive effect on grades.

The Role of “Objectivity”

Grading well is difficult, and it is especially difficult to sustain a wide range of grades absent reliable and replicable grading criteria (Johnson, 2003). Students cannot easily quarrel with a determination that they failed to differentiate an exponential function or to reproduce a chemical formula on the midterm. (Note that in this regard the fourth-semester languages courses are like math and science—one either does or does not correctly conjugate verbs or know specific elements of vocabulary.) The grading of English composition, on the other hand, intrinsically provides students with more opportunities to argue about the putative fairness of grades, thereby imposing costs on the instructor.

In this context, departments with a style of grading that involves interpretation and discussion of the work may choose to signal the quality of their best students in early classes by giving a handful of A’s and even A+’s, along with encouraging commentary, while giving good grades to middling students as well. The English department *could* impose stringent grading standards in the required freshman composition course without losing enrollments, but there is no reason for it to do so. Instructors would incur considerable cost and essentially no benefit.

generalization in the text would hold. Biology changed its organization and its numbering of courses during the period under study, and we are simply not sure if we are measuring it well. It is also the case, per widely asserted anecdote, that most biology majors are not all that interested in the subject per se, but are majoring in it in the hope of getting into medical school. If so, this could lead to low morale and low performance even in upper-division courses. Similarly, it is our subjective experience that many economics majors are not very interested in academic economics, but are more vocationally oriented. This could be part of an explanation for low upper-division grades in economics.

The same pattern would appear to be true of the laboratory sections of the required science courses. For the recent period, the average grade in lab sections in lower-division physics courses was 3.55, slightly to the A– side of A–/B+. The associated lecture courses, meanwhile, averaged 2.82, well below a straight B. Indeed, most of the grade inflation in introductory physics over the period of our analysis is due to increases in the grades in lab sections. The same group of labs that averaged a grade of 3.55 in 2005–2007 averaged 3.05 just 13 years earlier. If chemistry ever graded its labs and lectures similarly, it underwent its significant grade inflation in lab courses prior to the period analyzed here, as we observe large differences between average grades in labs and lectures throughout the period. In the later period, the introductory labs average 3.23, while the associated lecture courses average 2.89. It might be argued that labs can be graded objectively—full credit goes to those few souls whose labs actually come out right in every detail, and points are deducted for each component of the experiments or observations that is missed. The problem is that outcomes of lab assignments are notoriously unsuccessful so a grading standard based on success is as likely to reward luck almost as much as understanding or skill. Everyone seems to be happier in an environment where labs are graded principally by checklist—did the student do all of the required tasks and keep good notes?—rather than by success or failure. Laboratory experiments in electricity and magnetism seem to be more like book reports and less like calculus exams than we might have thought, and grading them lightly reduces the grief experienced by both students and faculty. Indeed, this understates the differences in grading standards. The introductory physics labs have *higher* grades, on average, than English courses in either the upper or lower divisions.

All of the fields that have low grades in required introductory courses have available to them “objective” grading methods in the sense that we mean here—it’s easy to construct examinations for which no one can quibble (at least not with much hope of success) about whether answers are right or wrong. We infer that two conditions are necessary to support a regime of relatively low grades: 1) grade-inelastic demand, which perforce exists for courses that are widely required; and 2) low hassle associated with giving low grades, which obtains when there are widely-agreed upon “objective” grading methods.⁴ The two factors together make it inexpensive to sustain low relative grades, and may have led to the bifurcation in the grading distribution across departments that began in the 1960s, as documented by Sabot and Wakeman-Linn (1991).⁵ We speculate that the practice of giving low grades in required courses often leads to lower grades in upper-division

⁴ We do not want to argue that quantitative fields must grade objectively and humanistic fields subjectively. It would be perfectly possible for, say, humanities to ask multiple-choice questions about the name of Sir Gawain’s horse, and for physics professors to ask for discursive essays about electricity and magnetism. The differences in the form of evaluation are not dictated solely by the differences in the nature of the subject matter, but also by norms and traditions in the different fields. That said, the relevant norms seem to be quite firmly established.

⁵ Sabot and Wakeman-Linn (1991) state that at Williams in 1962–3, with the exception of math, “there was little difference across departments either in mean grades or in the distribution of grades,” and they report

courses as well because departmental faculty and students are socialized to be comfortable with a grading regime that produces C's as well as A's and B's, even if the technology of grading in the upper division relies less on the relatively straightforward and simple instruments that are typically employed in the introductory courses.

Managing Departmental Enrollments

Departments that are subject to unwanted congestion could choose to reduce their relative grades, while departments that are teaching so few students that they are faced with losses in faculty strength may engage in a variety of strategies, including an increase in relative grades, to attract more students.⁶ However, this line of argument is difficult to test directly, in part because the relationship between student demand and administration behavior is typically slow and weak, and in part because it is difficult to find measures of changes in student demand that can be used to identify a supply response. We were unable to discern any systematic relationship in our Michigan data after trying all manner of lags between changes in enrollment and changes in average grades.

There are some highly suggestive examples, however. Over the time period of our Michigan data, grades fall in both the upper and lower division in communication studies, a field that has been highly congested and that has undertaken a number of policies—including reducing the range of courses offered and requiring successful performance in four lower-division courses for admission to the major—to reduce the number of majors. Enrollments grew markedly in statistics, in part as a result of other departments imposing requirements for work in the field, and again, grades seem to have fallen in response to the increased demand and congestion. At the same time, grades rose in a number of lower-division geology courses that are specifically aimed at students looking to fill their distribution requirements in the sciences without taking technically demanding introductory courses that are aimed at science majors.

Discussion and Policy Implications

Given faculty tastes for and interest in grading, one would expect that faculties would respond to economic incentives—to costs and benefits associated with

a similar finding for the other colleges and universities that they studied in the early 1960s, with the occasional exception of math and chemistry.

⁶ Notice that in either case, the department's interest may well diverge from that of an individual faculty member and that there may be variation in individual interests as well. A tenured faculty member who grades hard may be delighted with small enrollments of strong students, even if the department is at risk for losing a position somewhere down the road. A faculty member up for promotion in an environment where good student evaluations are essential to a successful promotion case may grade easy even if there is congestion in the department's courses or even in that faculty member's courses. How such conflicts are resolved (if they are) is surely highly variable, although in most fields and departments, individual faculty seem to have a good deal of autonomy in how they grade.

different grading regimes. The Michigan data we use here provide considerable support for such a proposition, assuming further that it is costly to departments to be either too large or too small, and that more students prefer easy grading courses to harder ones than the other way around.

Our evidence is consistent with the theory that two conditions are necessary to sustain a regime of relatively low average grades: 1) There must be enough enrollment demand so that the department is contributing sufficiently (as determined by the dean, typically) to the economy of the college. 2) Students must be limited in their ability to impose time costs and psychological costs on faculty through complaints about unfair grading. The easiest way to meet the second condition is through the use of “objective” assessment mechanisms that use right and wrong answers that are difficult to argue about, along with general agreement that providing the right answers is a reasonable test of mastery in the field.

Our principal empirical finding is that all of the lowest-grading departments at Michigan teach courses that are required and that are amenable to grading mechanisms that are hard to quarrel with. In most cases, this finding aligns with the general notion that math, the sciences, and economics grade hard, and the humanities grade easy, but the exceptions to this general proposition are powerfully suggestive of the power of the economic interpretation that we suggest here. Required science courses where it is difficult to assess the quality of performance beyond a checklist of procedures—notably the labs associated with introductory chemistry and physics courses—grade very lightly, with an average grade above B+. Meanwhile the fourth-semester language courses in French and Spanish—courses that provide a relatively low marginal cost mechanism for many students to meet the language requirement—grade much like introductory science courses. At the same time, introductory science courses in fields that are not required grade much higher than the introductory courses that are necessary for pre-med or engineering curricula, and introductory language courses in areas where students are starting without previous high school preparation also grade high, with one exception. And required freshman composition, a course which uses interpretative grading methods, has an average grade of B+ even though the grade-elasticity of demand is surely very low.

The most persuasive pieces of this story are the high-grading science courses and the hard-grading languages. This is especially so in light of the fact that the same departments revert to type in other settings. Physics grades easy in the labs but hard in its lecture courses. French and Spanish grade hard in the fourth-semester language course but easy (close to A-) in their upper-division literature courses that are principally taken by majors.

Several elements of the preceding argument would benefit from additional evidence. One involves real measurement of how easy it is to grade hard. In our discussion and interpretation we basically treat this as a binary variable. Surely it is continuous, and we wonder how far we could push the argument. Could it, for example, explain all or some of the difference between relatively hard-grading philosophy and easier-grading linguistics or comparative literature? Similarly, we would like to have something more than anecdote and the strikingly high grades

themselves to support our interpretation of high grades in the lab sections of introductory science courses. We have been told consistently that it is widely seen as “unfair” to grade students on the outcomes of the labs rather than the inputs and that students complain. It would be useful to get independent verification of this claim. We would also like to explore more fully two elements of faculty tastes that seem plausible and that are consistent with the differences in grades between the upper and lower divisions: the extent to which faculty desire to dissuade poor students from doing advanced work in their fields and the extent to which faculty like to protect and reward students who show aptitude and interest in their subjects.

We have said relatively little about the ongoing grade inflation—about 0.01 of a grade point per year on average—that we find in the data. It’s not hard to sketch plausible sources of upward pressure on grades over time. If some faculty increase their grades modestly (perhaps to improve teaching evaluations or perhaps just to reduce the potential for time and energy spent in conflict with students), the increase in the average may well lead other faculty to follow suit. In addition, if students at other universities are subject to weaker grading standards, faculty may change their behavior to be fair in the context of graduate admissions. Indeed, we spoke with a chemistry professor who had stuck to the standards of his own undergraduate work for decades, but who came to notice that incoming graduate students at Michigan often had better grades than graduates of the department with similar knowledge and skill. All of these pressures push grades up; none down. It’s rather like the general upward pressure on wages and prices observed in modern mixed economies except that in this case no one is paying any attention to any analog to the money supply. Given that uniformity of grading norms across academic departments left the building at least 40 years ago, the puzzle is not why so many departments grade easy, but why so many continue to grade relatively hard, even if, as we suggest, they only do so when the cost is relatively low.

How much should we be concerned about differential grading standards, especially in light of the fact that we see them as being, at least in part, rational responses to incentives and practices in the academy that would seem difficult to change? The answer depends on the uses to which grades are put, and here we know much less than we would like. Even with highly variable norms across fields, grades can convey a good deal of information about performance; we venture that most observers are well aware that the grade distribution in English is more compressed than that in math. At the same time, an A still conveys a judgment of strong performance in either field, and grades continue to provide somewhat noisy but usable information about academic performance. But oddly, we don’t know much about how grades are used by whom. If a student’s grade point average is widely used for consequential purposes, differential grading standards might do a good deal of mischief. For example, one area that uses grade point average is the awarding of academic honors and membership in honorary societies such as Phi Beta Kappa. To the extent that such honors matter to students, directly or indirectly, differential grading standards are a problem.

However noisy a signal grade point averages may be, when grades are interpreted in the context of departmental norms for given courses of study, they still

convey some information about performance. Much as with monetary inflation, in the absence of surprises, the important information is conveyed by relative prices, which—against the background of continuing inflation—are quite stable. Of course, the fact that grading norms differ across departments within and across universities weakens the signal conveyed by grades in particular courses as well as the signal conveyed by grade point average. Everyone may know that math typically grades harder than psychology, but it seems unlikely that anyone has a very good sense of relative norms in, say, journalism and comparative literature. What no one who is an intelligent user of grades can believe anymore is that the grading norms are identical across fields.

If grade distributions were to change abruptly, the result would be a reduction in the quality of the signals provided by grades. Suppose that math suddenly graded like English. A cohort of students would be surprised and a cohort of graduate admissions committees might make bad decisions. If English suddenly graded like math, the errors would be similar in kind, although opposite in sign. Moreover, if grades suddenly toughened in certain departments, the affected cohort of students would complain to their instructors, to the department, and quite likely to the dean (as described in Roosevelt, 2009).

Although concern over grade inflation has often focused on the weakening of signals, an equally or even more serious difficulty may be that, to the extent that students shop for grades, the purposes of liberal education are compromised. At least some students are responsive to grades, which implies that differential grade inflation has effects at every level—on the selection of introductory courses, on the probability of taking more courses in a field, and on the probability of majoring in a field. Such behavior may even affect overall career patterns if some students with an aptitude for quantitative work are scared off by low average grades in introductory math or chemistry. But even if the effect on careers is small, if differential grading standards lead students away from introductory courses in hard-grading fields and towards those that grade easier, students' overall educations will have been compromised, and for no purpose.

■ *Achen acknowledges support from an NIA training grant to the Population Studies Center at the University of Michigan (T32 AG000221).*

References

- Johnson, Valen.** 2003. *Grade Inflation: A Crisis in College Education*. Springer.
- Roosevelt, Max.** 2009. "Student Expectations Seen as Causing Grade Disputes." *New York Times*, February 17. http://www.nytimes.com/2009/02/18/education/18college.html?_r=2&scp=1&sq=grade%20inflation&st=cse.
- Richard Sabot, and John Wakeman-Linn.** 1991. "Grade Inflation and Course Choice." *Journal of Economic Perspectives*, 5(1): 159–70.

This article has been cited by:

1. Timothy M Diette, Manu Raghav. 2016. Class Size Matters: Heterogeneous Effects of Larger Classes on College Student Learning. *Eastern Economic Journal* **41**, 273-283. [[CrossRef](#)]
2. M. Ver Ploeg, P. Dutko, V. Breneman. 2015. Measuring Food Access and Food Deserts for Policy Purposes. *Applied Economic Perspectives and Policy* **37**, 205-225. [[CrossRef](#)]
3. E. Stockhammer. 2015. Rising inequality as a cause of the present crisis. *Cambridge Journal of Economics* **39**, 935-958. [[CrossRef](#)]
4. A. V. Thakor. 2015. The Financial Crisis of 2007-2009: Why Did It Happen and What Did We Learn?. *Review of Corporate Finance Studies* . [[CrossRef](#)]
5. L. Opazo, C. Raddatz, S. L. Schmukler. 2015. Institutional Investors and Long-Term Investment: Evidence from Chile. *The World Bank Economic Review* . [[CrossRef](#)]
6. A. Sunderam. 2015. Money Creation and the Shadow Banking System. *Review of Financial Studies* **28**, 939-977. [[CrossRef](#)]
7. B. G. Carruthers. 2015. Financialization and the institutional foundations of the new capitalism. *Socio-Economic Review* **13**, 379-398. [[CrossRef](#)]
8. Belayet Hossain, Panagiotis Tsigaris. 2015. Are grade expectations rational? A classroom experiment. *Education Economics* **23**, 199-212. [[CrossRef](#)]
9. V. Ioannidou, S. Ongena, J.-L. Peydro. 2015. Monetary Policy, Risk-Taking, and Pricing: Evidence from a Quasi-Natural Experiment. *Review of Finance* **19**, 95-144. [[CrossRef](#)]
10. N. Hautsch, J. Schaumburg, M. Schienle. 2015. Financial Network Systemic Risk Contributions. *Review of Finance* **19**, 685-738. [[CrossRef](#)]
11. Joe Hirschberg, Jenny Lye. 2015. The influence of student experiences on post-graduation surveys. *Assessment & Evaluation in Higher Education* 1-21. [[CrossRef](#)]
12. Ray Bachan. 2015. Grade inflation in UK higher education. *Studies in Higher Education* 1. [[CrossRef](#)]
13. Horacio Matos-Díaz, David McMillan. 2014. Measuring grade inflation and grade divergence accounting for student quality. *Cogent Economics & Finance* **2**, 915756. [[CrossRef](#)]
14. Michael A. Bailey, Jeffrey S. Rosenthal, Albert H. Yoon. 2014. Grades and incentives: assessing competing grade point average measures and postgraduate outcomes. *Studies in Higher Education* 1-15. [[CrossRef](#)]
15. X. Vives. 2014. Strategic Complementarity, Fragility, and Regulation. *Review of Financial Studies* **27**, 3547-3592. [[CrossRef](#)]
16. A. Monfort, J.-P. Renne. 2014. Decomposing Euro-Area Sovereign Spreads: Credit and Liquidity Risks. *Review of Finance* **18**, 2103-2151. [[CrossRef](#)]
17. N. Fligstein, J. Habinek. 2014. Sucker punched by the invisible hand: the world financial markets and the globalization of the US mortgage crisis. *Socio-Economic Review* **12**, 637-665. [[CrossRef](#)]
18. Kristin F. Butcher, Patrick J. McEwan, Akila Weerapana. 2014. The Effects of an Anti-Grade Inflation Policy at Wellesley College. *Journal of Economic Perspectives* **28**:3, 189-204. [[Abstract](#)] [[View PDF article](#)] [[PDF with links](#)]
19. I. Darnhofer. 2014. Resilience and why it matters for farm management. *European Review of Agricultural Economics* **41**, 461-484. [[CrossRef](#)]
20. R. C. Castro Miranda, S. R. Stancato de Souza, T. C. Silva, B. M. Tabak. 2014. Connectivity and systemic risk in the Brazilian national payments system. *Journal of Complex Networks* . [[CrossRef](#)]
21. G. Berg, K. Kirschenmann. 2014. Funding Versus Real Economy Shock: The Impact of the 2007-09 Crisis on Small Firms' Credit Availability. *Review of Finance* . [[CrossRef](#)]
22. D. B. Choi. 2014. Heterogeneity and Stability: Bolster the Strong, Not the Weak. *Review of Financial Studies* **27**, 1830-1867. [[CrossRef](#)]

23. G. Liang, E. Lutkebohmert, Y. Xiao. 2014. A Multiperiod Bank Run Model for Liquidity Risk. *Review of Finance* **18**, 803-842. [[CrossRef](#)]
24. I. Erel, T. Nadauld, R. M. Stulz. 2014. Why Did Holdings of Highly Rated Securitization Tranches Differ So Much across Banks?. *Review of Financial Studies* **27**, 404-453. [[CrossRef](#)]
25. R. Iyer, J.-L. Peydro, S. da-Rocha-Lopes, A. Schoar. 2014. Interbank Liquidity Crunch and the Firm Credit Crunch: Evidence from the 2007-2009 Crisis. *Review of Financial Studies* **27**, 347-372. [[CrossRef](#)]
26. G. Chabakauri. 2013. Dynamic Equilibrium with Two Stocks, Heterogeneous Investors, and Portfolio Constraints. *Review of Financial Studies* **26**, 3104-3141. [[CrossRef](#)]
27. G. Kyureghian, R. M. Nayga. 2013. Food Store Access, Availability, and Choice when Purchasing Fruits and Vegetables. *American Journal of Agricultural Economics* **95**, 1280-1286. [[CrossRef](#)]
28. S. Bougheas. 2013. Pooling, tranching, and credit expansion. *Oxford Economic Papers* . [[CrossRef](#)]
29. M. Kacperczyk, P. Schnabl. 2013. How Safe Are Money Market Funds?. *The Quarterly Journal of Economics* **128**, 1073-1122. [[CrossRef](#)]
30. C. Zhen, E. A. Finkelstein, J. M. Nonnemaker, S. A. Karns, J. E. Todd. 2013. Predicting the Effects of Sugar-Sweetened Beverage Taxes on Food and Beverage Demand in a Large Demand System. *American Journal of Agricultural Economics* . [[CrossRef](#)]
31. B. G. Carruthers. 2013. From uncertainty toward risk: the case of credit ratings. *Socio-Economic Review* **11**, 525-551. [[CrossRef](#)]
32. G. Kyureghian, R. M. Nayga, S. Bhattacharya. 2013. The Effect of Food Store Access and Income on Household Purchases of Fruits and Vegetables: A Mixed Effects Analysis. *Applied Economic Perspectives and Policy* **35**, 69-88. [[CrossRef](#)]
33. M. Andrews, R. Bhatta, M. V. Ploeg. 2013. An Alternative to Developing Stores in Food Deserts: Can Changes in SNAP Benefits Make a Difference?. *Applied Economic Perspectives and Policy* **35**, 150-170. [[CrossRef](#)]
34. Liam Delaney, Colm Harmon, Martin Ryan. 2013. The role of noncognitive traits in undergraduate study behaviours. *Economics of Education Review* **32**, 181-195. [[CrossRef](#)]
35. L. Le Maux, L. Scialom. 2013. Central banks and financial stability: rediscovering the lender-of-last-resort practice in a finance economy. *Cambridge Journal of Economics* **37**, 1-16. [[CrossRef](#)]
36. Maria De Paola, Francesca Gioia. 2012. RISK AVERSION AND FIELD OF STUDY CHOICE: THE ROLE OF INDIVIDUAL ABILITY. *Bulletin of Economic Research* **64**:10.1111/boer.2012.64.issue-s1, s193-s209. [[CrossRef](#)]
37. N. Arinaminpathy, S. Kapadia, R. M. May. 2012. Size and complexity in model financial systems. *Proceedings of the National Academy of Sciences* **109**, 18338-18343. [[CrossRef](#)]
38. Philip H. Brown, Nicholas Van Niel. 2012. Alternative class ranks using z -scores. *Assessment & Evaluation in Higher Education* **37**, 889-905. [[CrossRef](#)]
39. S. Nagel. 2012. Evaporating Liquidity. *Review of Financial Studies* **25**, 2005-2039. [[CrossRef](#)]
40. Z. He, W. Xiong. 2012. Dynamic Debt Runs. *Review of Financial Studies* **25**, 1799-1843. [[CrossRef](#)]
41. I.-H. Cheng, K. Milbradt. 2012. The Hazards of Debt: Rollover Freezes, Incentives, and Bailouts. *Review of Financial Studies* **25**, 1070-1110. [[CrossRef](#)]
42. Talia Bar, Vrinda Kadiyali, Asaf Zussman. 2012. Putting Grades in Context. *Journal of Labor Economics* **30**, 445-478. [[CrossRef](#)]
43. P. Feldhutter. 2012. The Same Bond at Different Prices: Identifying Search Frictions and Selling Pressures. *Review of Financial Studies* **25**, 1155-1206. [[CrossRef](#)]
44. Talia Bar,, Asaf Zussman. 2012. Partisan Grading. *American Economic Journal: Applied Economics* **4**:1, 30-48. [[Abstract](#)] [[View PDF article](#)] [[PDF with links](#)]

45. Rebecca Summary, William L. Weber. 2011. Grade inflation or productivity growth? An analysis of changing grade distributions at a regional university. *Journal of Productivity Analysis* . [[CrossRef](#)]
46. L. A. Bebchuk, I. Goldstein. 2011. Self-fulfilling Credit Market Freezes. *Review of Financial Studies* **24**, 3519-3555. [[CrossRef](#)]
47. F. Allen, E. Carletti, A. Leonello. 2011. Deposit insurance and risk taking. *Oxford Review of Economic Policy* **27**, 464-478. [[CrossRef](#)]
48. C. Gundersen, B. Kreider, J. Pepper. 2011. The Economics of Food Insecurity in the United States. *Applied Economic Perspectives and Policy* . [[CrossRef](#)]
49. N. Beale, D. G. Rand, H. Battey, K. Croxson, R. M. May, M. A. Nowak. 2011. Individual versus systemic risk and the Regulator's Dilemma. *Proceedings of the National Academy of Sciences* **108**, 12647-12652. [[CrossRef](#)]
50. A. Maddaloni, J.-L. Peydro. 2011. Bank Risk-taking, Securitization, Supervision, and Low Interest Rates: Evidence from the Euro-area and the U.S. Lending Standards. *Review of Financial Studies* **24**, 2121-2165. [[CrossRef](#)]
51. J. L. Lusk, K. Brooks. 2011. Who Participates in Household Scanning Panels?. *American Journal of Agricultural Economics* **93**, 226-240. [[CrossRef](#)]
52. A. Purnanandam. 2010. Originate-to-distribute Model and the Subprime Mortgage Crisis. *Review of Financial Studies* . [[CrossRef](#)]
53. C. Borio. 2010. Ten propositions about liquidity crises. *CESifo Economic Studies* **56**, 70-95. [[CrossRef](#)]
54. A. Demirguc-Kunt, L. Serven. 2010. Are All the Sacred Cows Dead? Implications of the Financial Crisis for Macro- and Financial Policies. *The World Bank Research Observer* **25**, 91-124. [[CrossRef](#)]
55. G. Caprio, A. Demirguc-Kunt, E. J. Kane. 2010. The 2007 Meltdown in Structured Securitization: Searching for Lessons, not Scapegoats. *The World Bank Research Observer* **25**, 125-155. [[CrossRef](#)]