

GRADING QUESTIONS AND ANSWERS

WHY ARE WE TRYING TO DO SOMETHING ABOUT GRADING?

1. Why do we grade our students?

- Like most of our peer institutions, Princeton has traditionally graded the academic work of undergraduates. The primary purpose of grading is educational: giving students accurate signals about the quality of their work helps them to calibrate the effectiveness of their efforts and motivates them to stretch to do their best work.
- Grading helps us to convey the achievements of our students to the external public who may admit them to graduate and professional schools, award fellowships to them, and employ them.
- Grading enables us to identify students who are not thriving in their programs of study and who may require special assistance, or may need to change direction or take some time away from the University, in order to make further academic progress.

2. What do we grade when we grade our students?

Individual faculty members will certainly have their own answers, and we do not presume to substitute the judgment of the Committee or the wisdom of department chairs. In our discussions, however, we returned again and again to a central theme: grades serve, above all, to evaluate the quality of a student's performance. They are meant, in other words, to signify how well the student did, measured principally in relation to the faculty member's expectations for the specific assignment, the course, or the independent project. Whether the assessment is absolute – that is, in relation only to those expectations – or whether it is also relative (measured in relation to the performance of other students in the course) is a matter for the individual faculty member to determine. We stress performance or achievement rather than effort; that is, grades are meant to signify what the student actually accomplished, not how hard the student tried or how much the student improved or how much the student could be expected to achieve given previous preparation or disciplinary background.

3. Why is it important to address our institutional grading practices? Why do grade inflation and grade compression matter?

- Grading, properly done, is an educational tool that assists students in evaluating what they have learned, how well they have learned it, and where they need to invest additional effort.
- Grading done without careful calibration and discrimination is, if nothing else, uninformative and therefore not useful; at worst, it actively discourages students from rising to the challenge to do their best work.

- Students are entitled to a fair and reasonable assessment of the work they have done; there should be some correlation between performance and reward.
- It does students no favor to grade them in a way that fails adequately to differentiate routinely good from really outstanding performance. We need to do a better job of distinguishing the excellent from the competent and of holding students accountable for negligent, weak, and unacceptable performance.

Students themselves attest that thoughtful discrimination in grading has important educational benefits. At the high end, it encourages students' best efforts. As one woman observed, "If I get the same grade for my very best work that I get for work that is not my very best, I'm less motivated to try to stretch as far as I can." At the low end, more careful discrimination in grading gives students an accurate calibration of the effort that is required to thrive at Princeton. One man wrote, "I got a D in math and I deserved it.... [N]ow I try to try harder." A woman observed, "My most significant educational experience at Princeton has probably been the day that I received my final grades for my first term. I didn't do as well as I should have. This made me realize that I had to work harder and, in a sense, differently. It made me realize that I was no longer in high school and that I had to get accustomed to a different style of learning."

HOW WOULD THE PROPOSED PLAN WORK?

4. The plan is premised on a social compact among the faculty. What does that mean? Why would we ask the faculty to enter into such a compact?

Institutional efforts to address grade inflation and grade compression raise what a number of department chairs have rightly described as a collective action problem. No individual department has an incentive to act unilaterally to address grade inflation. Indeed, for individual departments, all the incentives go the other way. There is nothing to be gained (and, faculty reasonably believe, a lot to be lost) from being tarred as the toughest graders in the University. But if all departments agree to act in concert, the incentives change; faculty can cooperate across departments to bring grading under more reasonable control.

We seek not to prescribe a set of rules to be handed down to the faculty, but rather to engage the faculty in a social compact, where departments agree on a common set of standards in the interest of the institutional good. That compact would comprise two basic operating principles:

- Departments would agree to meet an institutional grading standard;
- Each department would determine how to meet that standard, taking into account the range, size, and level of the department's courses.

It is important to be as clear about what the compact does *not* entail as about what it does. It does not require every faculty member to grade the same way. It does not require that every course have the same grade distribution. Rather, it seeks to vest

maximum flexibility and room for judgment in each individual department, at the same time that it asks each department to agree to meet a common institutional standard.

This kind of compact, balancing the prerogatives of the individual faculty member and the needs of the individual department, on the one hand, and the good of the institution as a whole on the other, describes much of the way we do academic business at Princeton. For example, we set common standards for the number of weeks in the teaching semester and for the number of hours of class meetings in any course. We set common standards for the proposal and approval of new courses, so that the desires of the individual faculty member are considered, first, in the context of the overall needs of the department, and then vetted through a centralized approval process where the Committee on the Course of Study acts corporately on behalf of the faculty. Extending the balancing of individual prerogatives and the common good to grading seems to be a natural extension of the way we do our academic business as a faculty.

5. What targets should we set?

We propose to control the number of A (that is, A+, A, and A-) grades awarded in undergraduate courses and independent work. This approach has two advantages: it is simpler to understand and implement than any other plan we considered; and it allows maximum flexibility for departments to determine how to achieve the desired objective, with due regard for the level, size, and complexity of their course offerings. The expectation, of course, is that if there are limits on A's, other grades will fall into line.

The Committee on Examinations and Standing gave the most careful consideration to the right limits to set. With the strong encouragement of the department chairs, we chose 35 percent A grades in undergraduate courses and 55 percent A grades in independent work. (The higher limit for independent work reflects the intellectual maturity, focus, and expertise that many juniors and seniors bring to their independent work.) In selecting those limits, the Committee was influenced by two considerations: the limits needed to be achievable in the Princeton context, and the difference between the proposed limits and current practice needed to be large enough to be meaningful. As for the first consideration, we note that 35/55 is by no means outside the Princeton experience; it resembles very closely the grading patterns at Princeton in undergraduate courses and independent work in the period F87-S92, and it describes (or comes close to describing) current grading patterns in some departments. As for the second consideration, the Committee has been guided by the conviction that if we really mean to make a difference in tackling grade inflation, we should take a bigger rather than a smaller step. The Committee takes very seriously the conviction of the large majority of department chairs that the bolder we are, the more the outside world will notice what we are doing, and the less likely we will be to risk disadvantaging our students.

6. What about grades below the A range?

As we have already said, this plan is premised on the assumption that if we control A grades, other grades will fall into line in appropriate relationship to those A's. For the plan to work, it is essential that the remaining grades not all be bunched in the B range. Faculty are strongly exhorted to recognize adequate or acceptable work with grades in the C range and to use D's and F's as appropriate to denote weak to very poor performance.

Faculty are exhorted also to accompany grades with evaluative critical comments that enable students to understand the strengths and limitations of their work and that give them guidance about ways of improving their performance.

7. How should we understand the special challenges of small courses and small departments?

At the same time that we are striving to achieve a common grading standard, we need to be sensitive to the ways in which course formats may affect grading patterns. Large-enrollment courses are likely to afford more latitude for the fullest use of the grading scale. Very small courses that depend on close interaction between students and instructor are likely to present more of a challenge. The same is true of advanced courses that may enroll a very high proportion of departmental concentrators.

Departmental enrollments are also likely to have an effect on the relative ease of adherence to institutional grading limits. Very small enrollments may suggest a high degree of self-selection, either in course enrollments or among concentrators. We should be sensitive to the variability of experience with respect to the range of student performance in our largest vs. our smallest departments.

8. What tools are available to chairs to make this new plan work?

The proposal is grounded in the expectation that grading, while still properly a private, individual activity, is also a matter of shared departmental culture and concern. The proposal counts on the chair to lead departmental faculty in deliberating about how best to meet the stated grading limits, taking into account the range, size, and level of difficulty of undergraduate courses and the collective expectations of the faculty for independent work. That might mean agreeing on one standard for normal grading patterns in large introductory lecture courses, and a rather different standard for grading in small courses at a more advanced level. It might mean agreeing that everyone ought to try his/her best to hit the same target, with allowances for slippage in one direction or another depending on student performance. The strategy for achieving the overall target would be a matter for collective deliberation. We distributed model grading patterns to chairs this fall (and can do so again) to provide some guidance about approaches that could work.

Several chairs have already suggested how they and their departments would be likely to proceed. With respect to course grades, one approach is to have the faculty

deliberate collectively about ways of meeting the grading limits in light of the particular mix of courses the department offers. Such deliberation would be informed by the department's own grading history as well as model grading patterns in other departments. A second approach is to ask everyone to try to meet the same standard. Gene Grossman (Economics) explains how he might proceed: "Suppose the limit on A's is 35%. I could envision instructing my colleagues that they should ordinarily submit a grade sheet with no more than 1/3 A's. I would leave them the flexibility to go above 1/3, but any such grade sheet would need to be accompanied by a note to me explaining why the faculty member felt it was necessary and appropriate in the particular case. Any faculty member who regularly submitted such notes would be invited for a discussion."

As for independent work, a number of departments have already tried different ways of controlling senior thesis grades. One approach is collective deliberation before final grades are assigned. Several of the science departments follow such a method. Adviser and second reader evaluations are presented, along with proposed grades; all faculty then discuss the merits of each case and come to agreement on final grades. Martha Himmelfarb (current chair) and Jeffrey Stout (acting chair last year) explain the practice in Religion. Some years ago, Himmelfarb notes, the faculty "instituted a practice of meeting together as a department to confirm grades for the thesis and senior comprehensive exam.... This allows faculty members to compare notes on grading standards, and while it can be tedious, I think it has made us the rare department that is tougher on independent work than on courses." Stout elaborates: "In each case, the two readers describe their student's performance and explain their reasons for proposing particular grades. When all of the cases have been discussed, we reconsider disputed and borderline cases in light of the general discussion. Only after concluding this rigorous process do we assign grades."

Another approach is to specify in detail the criteria by which independent work will be evaluated. Molecular Biology and Chemical Engineering have developed extremely detailed parameters for evaluating senior theses. The adviser and the other reader or readers evaluate the thesis by scoring the effectiveness of as many as nine or ten different aspects of each student's work. The Molecular Biology guidelines explain, "The scores on these parameters, which all receive equal weight, are first averaged for each grader. A thesis grade is then calculated, giving the adviser's grade a weight of 50 percent and each reader's grade a weight of 25 percent. All of the departmental theses are then ranked (laboratory theses are ranked separately from nonlab theses) and letter grades assigned based on University-wide average percentages-of-the-class that receive each grade. Individual thesis grade adjustments up or down can be made." The criteria are fine-tuned each year based on comments from the faculty.

In addition to leading their departments in instituting policies and practices of the kinds already described, chairs might offer their faculty colleagues suggestions about ways of making more discriminating judgments about the quality of student work. Thomas Espenshade (previous chair of Sociology) suggests that the various components of course work be given numerical grades; only at the end, when numerical grades have been averaged, are letter grades assigned. One can argue, of course, that making

discriminating judgments is relatively easy to do in more quantifiable disciplines, but more challenging to accomplish in disciplines where the work to be evaluated comes principally in the form of narrative essays, where “proof of inadequacy” is more difficult to demonstrate. In the latter case, Caryl Emerson (Slavic) suggests the possible utility of “‘anonymous submission’ of narrative work [i.e. interpretive essays],” where students submit papers identified only by numbers distributed by a member of the department’s office staff and unknown to the faculty member. That allows grading to be done “strictly on [the] merits” of the work, without taking into account other factors that may come into play in small, interactive classes (personality, performance in seminar, general intelligence, effort). She suggests also the advantages of incorporating in advanced literature courses “LOTS of grading variables: in addition to the final paper, two short fixed-question critiques, two ID quizzes (with definite right-and-wrong answers); attendance taken for every lecture and precept AND precept performance graded, so there is simply more to work with (some of it quite quantifiable).”

In response to the Committee on Examinations and Standing’s successive reports to the faculty, individual members of the faculty have written to offer suggestions of their own. William Howarth (English), for example, urges “blind grading of all papers, exams, and independent work. In blind grading, preceptors read the work of other preceptors’ students, and in double-blind reading, student names do not appear on the work, only numbers assigned by...administrators. Anonymity allows the faculty to be more objective and the students have fewer opportunities to pressure their teachers into awarding inflated grades.” Doubtless other faculty members will have suggestions about techniques that work well for them; as we receive them, we can supply chairs with an inventory of good suggestions drawn from a number of departments so that they can share them with their colleagues.

We appreciate that there may be a range of views among members of the faculty about the proposed grading standards. We were told repeatedly in our conversations with chairs that many faculty are looking for guidance about grading – indeed, will welcome a clear articulation of institutional grading standards. At the same time, we were also told that there are faculty who insist that grading is solely their own prerogative and resist the notion that “Nassau Hall” will tell them how to grade. (We note, as an aside, that “Nassau Hall” and “West College” are not trying to tell the faculty how to grade. These proposals come at the express instruction of the department chairs; if they are adopted, they will express not the will of the administration, but rather the collective vision of the faculty about grading at Princeton.)

In the end, in the event that a faculty member simply refuses to cooperate, the most powerful tools available to a chair will be moral suasion, jawboning, and the collective moral weight of the rest of the department – that is, of a group of faculty who have agreed to enter into a compact that ought not be disrupted by the unwillingness of one individual to do his or her part. If all of that fails, the Dean of the College and the Dean of the Faculty may be able to provide some help.

9. What tools are available to heads of large courses?

We have been asked more than once how heads of large courses can be expected to do their part in meeting departmental grading limits when individual preceptors grade their own precepts. The import of that question is that when large courses function like loose confederations of small courses, it will be difficult to make a lot of headway in achieving a reasonable distribution of grades.

We quote from the *Guide to Good Grading Practices*, which offers guidance about what heads of courses can do:

1. Establish practices to maintain equity in grading standards and work expectations across precepts and classes in multisection courses.
2. Develop a course definition for what constitutes work that would earn an A, B, C, D, or F. Discuss what particular criteria you will evaluate on specific assignments.
3. Circulate and discuss with your teaching staff representative A, B, C, D, and failing papers, assignments, or exam answers. Hold a collective grading session in which you and your teaching staff grade the same small set of papers or exams, and then discuss the outcomes in order to form a working consensus.
4. Establish and employ common procedures for accepting late work, penalizing grades, and grading nonattendance or nonparticipation. Work within established departmental standards.
5. Spot check the grading of the teaching assignments by the members of your teaching staff.
6. Encourage and monitor the critical feedback and evaluation provided by your teaching staff.
7. Grade final exams by question rather than by precept. In other words, have members of the teaching staff each grade a single question on the exam rather than grading the full exams of their own precept students. Such a practice ensures consistency in grading on questions and precludes bias in the final exam grading.
8. Review the grades from each precept and monitor the proposed grade distributions across precepts or classes before submitting final grade sheets to the registrar.

IS THE PLAN PRACTICAL? ACHIEVABLE? HOW WOULD IT AFFECT OUR STUDENTS?

10. How would 35/55 position Princeton relative to peer institutions?

The Dean of the College asked colleagues at the seven other Ivies, Stanford, MIT, and Chicago what percentage of A grades were given in undergraduate courses at their respective institutions last year (that is, in 2002-03). The data are not absolutely precise; in at least two cases, institutions sent information that is two years old, and it is not entirely clear that everyone is counting in exactly the same way. That said, the numbers (rounded to the nearest whole) ranged from 44 to 55 percent A's. Between those two poles, two institutions fell around 45-46 percent, four (including Princeton) between 47 and 48 percent, and three in the 49-51 percent range. (We have no systematic way of comparing independent work grades.)

In short, 35 percent A's will set us apart from the pack in a way that identifies Princeton as a real leader in tackling this problem.

11. Can departments reasonably be expected to meet these targets?

Based on 2002-03 grading data, eight departments (one in the humanities, one in the social sciences, six in the natural sciences) fell between 35 and 39 percent A's in course grades; nine departments (three in the humanities, two in the social sciences, two in the natural sciences, two in engineering) fell below 55 percent for senior thesis grades; nine departments (four in the humanities, three in the social sciences, two in the natural sciences) – but not the same nine -- fell at 57 percent or below for junior independent work grades.

In short, the fact that some of our departments currently meet or come close to 35/55 suggests that these are achievable targets.

12. How would a new institutional grading standard affect the fortunes of our students in graduate and professional school admission, national fellowship competitions, and the job market?

We have tried to answer that question in two ways: first, by estimating the effect on students' grade point averages, and second, by asking for advice from professional school admission deans, administrators of national fellowship competitions, and major employers.

a. Estimating the effect on grade point averages

While it is difficult to figure out how to model the effect of 35/55 on any individual student's grade point average, we think we can estimate the overall effect on the mean grade point average in all undergraduate courses at Princeton. In our own history, as we noted earlier, we came closest to awarding 35 percent A's in undergraduate courses in the period F87-S92 (the actual percentage then was 36.7). The mean GPA in that five-year period was 3.21 – down from 3.36 in the most recent five-year period, F97-

S02. In the humanities, the mean GPA in undergraduate courses was 3.47 in F97-S02, 3.34 a decade earlier; in the social sciences, 3.32 and 3.18; in the natural sciences, 3.16 and 3.03; in engineering, 3.37 and 3.16.

It is reasonable to imagine, then, that the impact of 35/55 on any individual student's grade point average is likely to be small.

b. Estimating the impact on graduate and professional school admission, fellowship competitions, and employment

At the behest of the department chairs, the Dean of the College has spoken to or corresponded with deans and directors of admission of leading law schools and medical schools, administrators of national fellowship competitions, and recruiters from employers with whom we do a fair amount of business.

Some excerpts from correspondence (quoted) and conversations (paraphrased) with the medical schools:

“The School of Medicine applauds your proposal, and wishes we could see more universities make a concerted effort in that same vein. Once you alert us to the change, we will be certain to review applications from Princeton with the knowledge that we are seeing appropriate grades for course work accomplished.”

“The medical schools would welcome a serious attempt by Princeton to control grade inflation. If Princeton makes a clear statement that this is afoot, I think the schools would interpret Princeton grades accordingly.... Other colleges might follow your example – a salutary influence.”

“We have learned by experience about schools with tougher or more lenient grading policies and have reviewed student performance accordingly. If there would be conscious change at Princeton [made known to us], we would also act accordingly without hurting your students.”

It is wonderful that Princeton is willing to address grade inflation. The medical school is certainly able to judge grades in the context in which they are awarded. Princeton should explain its new grading practices and include a grade distribution grid with committee letters of recommendation. As long as Princeton shows the medical school what the grades mean, more rigorous grading would absolutely not hurt Princeton applicants. Indeed, that medical school would be more flexible in evaluating Princeton applicants if they knew that Princeton grades had real meaning.

Conversations with the law schools yielded the following observations (paraphrased):

Grade inflation is a pervasive problem (and Princeton is not the worst of the offenders). It would be terrific if a major institution were courageous enough to take some leadership to address it.

It is a real challenge now for law schools to pick out the really outstanding 3.7/3.8's. If Princeton adopted a more rigorous grading standard, it would very likely help stronger students, since the schools would understand that a 3.8 from Princeton is a real 3.8, unlike inflated 3.8's at other institutions. It could also help excellent students from Princeton who may test less well on the LSATs by distinguishing them in comparison with applicants from other places.

If Princeton educates the law schools about the new grading standard, they will be able to evaluate Princeton applicants in the context of that standard. Princeton can accomplish that education by being very clear about the new standard – by sending a letter along with the transcript, even considering annotating the transcript, at least for a transitional period until the schools become accustomed to the new practices.

The law schools already have experience admitting students from a handful of first-rate institutions that have held the line on grades; they do not pass up those students because their GPAs are lower than others, and there is no reason to think that they would behave any differently with respect to Princeton students.

When a student from Princeton (the same is true of any other college or university) applies to law school, the Law School Admissions Council sends the law school a report showing how that applicant's LSAT score and GPA fit in the context of a running average of LSAT scores and GPAs for all Princeton students who have applied to any law school over the past three years. In other words, the law schools have a kind of "imputed rank in class," a ready-made tool to understand each Princeton applicant in the Princeton context. It will be essential for Princeton to educate LSAC about the new grading standard. The period of transition from the old to the new grading standard will require some special attention.

The best schools always look beyond GPAs. What kinds of courses has the student taken? How rigorous are they? What do recommenders say about the applicant's quality of mind? originality? analytic rigor?

The top-tier law schools strongly encourage us to proceed with our plan. At the same time, however, they point out that we could incur some risks at some second- and third-tier schools. One issue is sensitivity to US News & World Report rankings of law schools; the mean GPA of the admitted class gets reported (including 25th and 75th percentiles) and affects a school's standing. For schools that are especially concerned to better their rankings, there could be some effect at the margins (e.g., a school thinking about taking a Princeton student with a 3.3 GPA off the waitlist in the summer could worry about what such an admit would

do to the mean GPA of the entering class). A second issue of possible concern: some lower-ranking schools with large numbers of applications employ algorithms (or formulas) with certain weights given to LSAT scores and to GPAs (the algorithms differ from school to school). The algorithm produces a number; all candidates above a certain number may be admitted, all candidates below a certain number may be rejected, and those in the middle may be evaluated on a case-by-case basis. If the algorithms are not sufficiently sensitive to adjust for our new grading standard, we could, again, see some difficulty at the margins.

In sum (and here we interpret the conversations): Princeton students do very well at the leading law schools; there is no doubt that they will continue to want – and admit – Princeton students. Most schools say that if we reform our grading standards, we will not compromise the fortunes of our applicants.

Next, some excerpts from correspondence with administrators of national fellowship competitions (quoted):

From a competition for scholarships in the United Kingdom: “We would welcome such a change. In fact, if communicated clearly (e.g. with tables showing actual grade distribution), I think it would help your best students. And, of course, I hope your example would be followed. National publicity for the change would be likely, I assume, further reducing the likelihood of any harm to your students. As Princeton, you have the ability to do this without cost to your students in a way that most other institutions would not.”

This same administrator continues: “This isn’t to say that it is possible that some average Princeton students have been advantaged by grade inflation and would be less likely to be so advantaged in the future. The tradeoff is greater advantage to your truly outstanding against possible disadvantage to your average. Frankly, however, I think the risk of this is relatively small in a competition like [ours] which relies so much on multiple references and interviews. The average student with a great transcript is usually apparent given our screens.”

And, further, “...If you make the change, we would expect even your very best to get some Bs and perhaps a C or two in very challenging courses outside their concentration, and that should not harm their chances.”

This administrator wishes us “good luck,” noting that “this needs to happen,” and concludes: “We want to identify the very best. A new grading system will help us do that.”

From a competition for scholarships for study in Ireland: “I...certainly applaud Princeton’s initiative. In seeking to erode grade inflation by limiting the number of top grades, then it will be important to attach a covering note explaining the grade distribution policy because, obviously, a B in that system might be the equivalent of an A at many other schools. Actually, this is precisely why we do

not have a minimum GPA...for our Scholarship since we recognize that each school and each student's curriculum are different. We look at the individual applicant carefully in the context of both the institution and the academic program.”

From a competition for fellowships to support graduate study in the humanities: “Criteria for [our fellowships] include the following: personal statement, essay sample, three faculty recommendations, language background, GRE scores, transcript with grade point, brief discussions of why student is applying to programs at various universities. Semifinalists are also given a half-hour-in-person interview. Given that range of information, the grade point does not have a large significance unless it is stunningly low.... In all, I do think you are wise in attaching a statement of grading policy.... But otherwise, I think any fear that more realistic grades would hurt a student's opportunities is completely unfounded. Instead, it will make a high grade point newly meaningful.”

From a competition for scholarships for study in Germany: “I have contacted a number of colleagues in my organization for advice. Everybody was very impressed with the courageous readiness of Princeton University to address problems of grade inflation.”

This same administrator continues: “We do think – given appropriate information – we will be fully willing and prepared to take account of different grading standards at different US institutions. ...such differences in grading standards are a well known and accepted fact in [our national] context....”

From a foundation providing grants for study in the applied physical sciences: “My guess is that your change in grading procedures will cause those truly gifted persons to emerge from the overly compressed, high grade point ‘noise.’ We also are very aware of the quality of the teaching and grading at those schools from which most of our applicants arise (e.g., Princeton), hence a change of the nature that you propose will be noted as a positive event.”

From a foundation providing grants for graduate or professional study for new Americans: “We look beyond grades and overall GPA when we award [our] Fellowships. The transcript, plus your explanation, would be sufficient. Never comparing finalists by grades in the first place, we would continue to have confidence in students and graduates from Princeton.”

Finally, some excerpts from correspondence with employers (quoted):

From a leading consulting firm: “A change in grading system is not a problem, from our standpoint. In fact, a reduction in grade inflation would make it easier

for companies such as ours to distinguish between truly outstanding students and those who are simply above average or very good.”

From a major financial services firm: “Princeton is one of the most important schools that we recruit from.... [Our firm] only recruits on campus at 20 schools.... Princeton is also one of the few schools where 5 divisions of the company recruit...usually two divisions recruit on campus at our other core 20 schools. Also...we have a very strong Princeton alumni group that is committed to recruiting and nurturing the careers of Princeton students who are interested in finance. Although I am aware that the entire student body is comprised of high caliber students, the pool of applicants that we have received from Princeton in recent years has been exceptional. That all being said, in my opinion, if Princeton changed its grading policies, I do not think that would damage the students’ relationship with [our firm] at all. We would be willing to work with the Career Center and other members of the Princeton community to understand the new grading policies and adapt to the changes.”

From a corporate strategic planning group: “We look at Princeton students in comparison to other Princeton students in determining first-round interview candidates, but we compare them to their peers at...other institutions in determining final-round interviews and, ultimately, offers. Generally GPA isn’t a pivotal issue to us in terms of making offers, but I can’t say that it’s not a factor at all. And, since no obvious public stance has been taken by any of these universities that their grading system is different from that of their peers, I do believe that we tend to view students as having done the same level of work if they were to achieve a 3.5 GPA at Princeton or a 3.5 GPA at [comparable] schools.... Assuming you could make this a very public change, it may be the case that in the future we would be indifferent between a Princeton student with a 3.2 and a [student from a peer institution] with a 3.5, but I can’t say it’s true today.”

WHAT IS THE RELATIONSHIP BETWEEN GRADES AND COURSE EVALUATIONS?

We often hear concerns about the presumed pernicious relationship between grades and course evaluations. The argument is that our evaluation system fuels grade inflation: faculty (especially junior faculty) and assistants-in-instruction feel pressured to give high grades in order to insure that they will receive favorable evaluations from their students. At the same time, the sequencing of evaluations (usually submitted in the last week of classes) and final grades (usually made known some weeks later) argues against such a causal relationship, and we have ample anecdotal evidence to support the contrary case: we know that some of our most popular teachers are among our toughest graders.

In order to make sense of this issue, we asked Jed Marsh, Vice Provost for Institutional Research, to analyze the relationship, if any, between faculty grading and

student course evaluations. He compared mean course evaluation scores and mean course GPA for classes, seminars, and lecture courses offered between fall term 1999 and spring term 2003 that met the following criteria:

- The course offering had four or more students
- Four or more evaluations were submitted by students in the course
- The instructor of record had been an instructor of record in at least three other courses during this period
- The number of evaluations submitted did not exceed the number of enrolled students

In all, he analyzed 4,850 course offerings and found that on average, a one point increase in the mean GPA is associated with an increase of more than 0.7 in the teaching evaluations. While this relationship is statistically significant, the individual data points do scatter considerably about this trend. In addition, we do not have sufficient evidence to determine whether course evaluations drive grade inflation. To the extent that that may be the case – if it is at all – the new grading standard will serve as a countervailing pressure. That is, instructors who may imagine that their continued employment will be aided by positive evaluations (with those, in turn, aided by liberal grading) will now also understand that responsible grading is an important component of teaching performance.