

Report from the Ad Hoc Committee to Review Policies Regarding Assessment and Grading

August 5, 2014

Contents

1 Introduction	2
1.1 Process	2
1.2 Summary of findings and recommendations	2
1.3 Organization of this report	3
2 Effectiveness of the current policy	3
2.1 Are the goals appropriate?	3
2.2 Does the policy achieve these goals?	4
3 Public perception of the current policy	10
3.1 Survey results	10
3.2 Views from the USG	13
4 Side effects of the current policy	13
4.1 Anxiety among the students	13
4.2 Competitiveness outside Princeton	14
4.3 Admissions yield	16
4.4 Impact on freshmen	17
5 Alternative policies considered	17
5.1 Change the currency	17
5.2 Intentionally inflate grades	18
5.3 Report more information on the transcript	18
5.4 Treat freshman year differently	19
A Committee members	20
B Charge to the committee	21
C Survey results	22

1 Introduction

In April 2004, the faculty adopted a grading policy that aimed to provide common grading standards across academic departments and to give students clear signals from their teachers about the difference between good work and their very best work. The policy recommended that, over time, each department award no more than 35% of A-range grades for course work and no more than 55% of A-range grades for junior and senior independent work.

In October 2013, President Eisgruber charged an ad hoc faculty committee with reviewing the University's policies for how student work is evaluated. The committee was composed of faculty members from each division and included current or past members of the Committee on Grading and the Committee on Examinations and Standing, as well as faculty who regularly teach large numbers of undergraduates. The members of the ad hoc committee are listed in Appendix A.

The president asked the members of the committee to explore whether the grading policy's objectives remain the appropriate ones against which to judge Princeton's assessment practices. He also asked them to examine whether the policy achieves the University's pedagogical goals effectively, with as few negative consequences as possible, or whether there are better ways to reach those goals. The president's charge to the ad hoc committee is attached to this report as Appendix B.

1.1 Process

In the course of its work, the committee sought broad input. We surveyed current faculty members and undergraduates, created a public comment website to elicit feedback from alumni and parents, and met with a range of administrators, faculty members, and students to discuss their perspectives on and experiences under the grading policy. In addition, the committee analyzed existing quantitative and qualitative data and consulted with colleagues at select peer institutions to better understand their policies and practices with regard to grading.

1.2 Summary of findings and recommendations

Based on what we learned, we recommend a number of changes to the current policy, which we summarize here.

Remove the numerical targets from the grading policy. Such targets are too often misinterpreted as quotas. They add a large element of stress to students' lives, making them feel as though they are competing for a limited resource of A grades.

Charge departments with developing their own grading standards. We recommend replacing the numerical guidelines with a set of grading standards developed and articulated by each department. These standards, which could be reported to the dean of the college as was recently done with standards for independent work, would help meet the goal of the current policy of maintaining "consistent standards across departments." Note that while "consistent standards" has been interpreted to mean "consistent grades," standards are not the same as grades, as we discuss in Section 2.1. Indeed, it is difficult even to define how to compare the "standards" among departments in vastly different fields. Furthermore, the data discussed in Section 2.2 show that the grading policy has had limited success at improving consistency

of grades across departments, and in some cases has made the disparity worse (for instance, between students concentrating in Engineering as opposed to Mathematics or Physics).

As we show in Section 2.2, grades began to decline a year before the grading policy was put in place. A reasonable interpretation is that this decline was due to conversations the dean of the college was having with departments about grade inflation, which resulted in an increase in general awareness of grading policy among the faculty. We recommend that the dean of the college continue to monitor grades across departments and communicate with departments when their own grading policies appear to be changing over time. We further recommend that departments internally review their grading history on a regular basis to ensure consistency with their own promulgated grading standards. We suggest that this sort of continuing conversation is a more effective and positive way of having an appropriate grading policy, and with fewer undesirable side effects, than setting numerical targets.

Dissolve the Committee on Grading. The emphasis should move away from “grades,” and instead focus on “quality of feedback.” To this end, we feel it would be appropriate to charge the newly-formed Council on Teaching and Learning with advancing efforts to improve quality of feedback, in place of the Committee on Grading’s current focus on limiting grade inflation.

No changes to the freshman year. While not part of our formal charge, we were also asked to examine whether grading in freshman year ought to be treated differently in order to help ease the transition to Princeton. After careful consideration, we do not recommend any changes to the grading in freshman year. As discussed in Section 5.4, we considered many possible scenarios, including covered grades, or no grades, for a semester or a year, and we think that all of these have the potential to create more problems than they solve. We do note that freshman year (especially the first semester) is a stressful time for students, particularly for first-generation college students and those from disadvantaged backgrounds. However, we are not convinced that grades are the central issue here. In order to address this issue, some other group could be convened to focus exclusively on freshman year and to consider more than just grades.

1.3 Organization of this report

The remainder of this report details the process and analysis we went through to reach our conclusions. In Section 2, we discuss whether the goals of the current policy are the appropriate ones and whether the policy has been effective in meeting those goals. In Section 3, we look at the public perception of the policy, discussing the results of the surveys mentioned above. Next, in Section 4, we discuss our understanding of the main side effects of the policy. We then briefly describe, in Section 5, some alternative policies we considered but did not end up recommending.

2 Effectiveness of the current policy

2.1 Are the goals appropriate?

The current grading policy has two main objectives: a) to maintain “fair and consistent” grading standards across academic departments; and b) to give students clear signals about the

quality of their work. We were concerned throughout our discussions to determine whether these goals are appropriate.

Our findings suggest that the first goal is not appropriate. In our online surveys, to be discussed in more detail in Section 3.1, a majority of both students (68%) and faculty (60%) who responded agree or strongly agree that “Students in one academic department should be graded according to the same standards as students in any other department.” But only 5% of students and only 6% of faculty actually believe the current policy is effective in meeting that objective. (Another 19% of students, and another 32% of faculty, believe the policy is “somewhat effective.”) In our view, this disparity arises because “consistent standards” is often interpreted, by students and instructors alike, to mean “consistent grades”—an understandable misconception given the emphasis in the current policy on maintaining (in effect, standardizing) specific percentages of A-range grades.

But standards are not the same as grades. Standards are the evaluative rubrics departments and instructors develop to convey to students the specific expectations for their work and against which that work will be graded. Grades measure the extent to which students meet those expectations. “Consistent standards” is an admirable goal in the abstract, but in practice it is our judgment that it is difficult if not impossible to compare standards across departments in different fields. Meaningful standards are—or should be—not only detailed but also highly course- and discipline-specific. They can only be easily compared across disparate fields if they are vague or generic. As a consequence it also seems to us less important that the same percentage of students in every department receive A-range grades—the current implication of the “consistent standards” objective—than that there be a high correlation between the grades students receive and the evaluative rubrics in the specific courses they take. Thus we have come to feel strongly that departments should spend their time developing clear and meaningful evaluative rubrics for work within their disciplines rather than aligning grades to meet specific numeric targets.

Our findings suggest, on the other hand, that the second objective of the current grading policy is appropriate: faculty should use grades to give students clear feedback about the quality of their work. This objective strikes us as integral to any grading policy. The online survey results affirm this belief: 79% of students and 93% of faculty who completed the survey agree or strongly agree that “Grades should give clear signals about the difference between ordinarily good work and the very best work.” At the same time, we do not believe that grades are the only means, or even necessarily the most effective means, of giving such feedback. Instead it is the quality and content of instructor feedback—how detailed, informative, and timely it is—that matters most and that should receive the highest emphasis in any grading policy.

2.2 Does the policy achieve these goals?

We begin with a summary of the history of grades at Princeton that led to the current policy, and of the experience since the current policy was implemented. The analyses below rely on data provided by the registrar on the course-specific distribution of grades at Princeton from academic years 1974–2013. It does not include information on grades in junior and senior independent work.

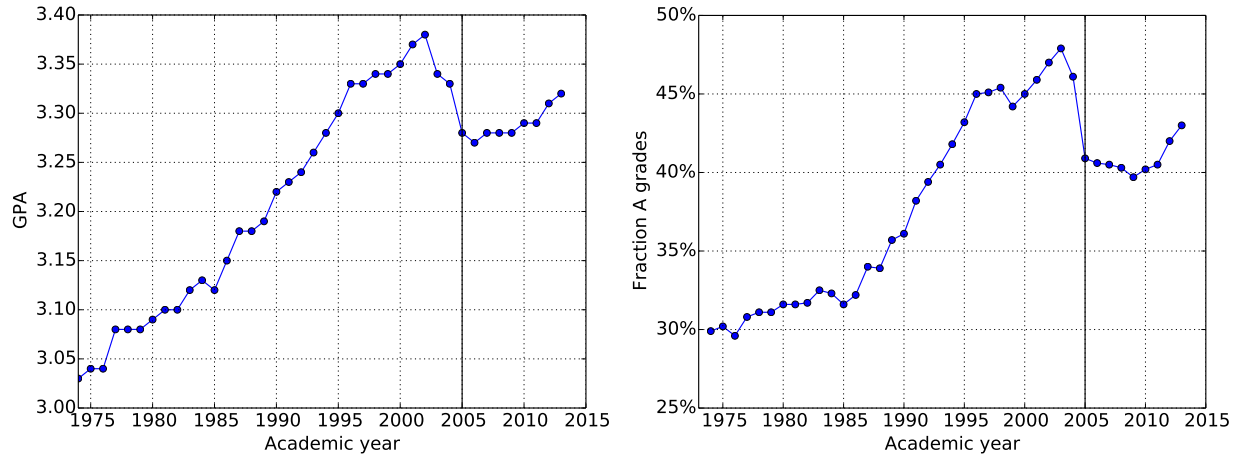


Figure 1: Average GPA and fraction A-range grades by year, 1974–2013.

Average GPA and the fraction of A-range grades

The left panel of Figure 1 shows the average GPA at Princeton each academic year from 1974 to 2013; the right panel shows the fraction of grades at Princeton in the A range (A+, A, A–) each year over the same period. Several interesting features are apparent from these figures:

- There was substantial grade inflation from 1974 through 2003. Over this period, the raw average grade at Princeton increased from 3.03 (B) to 3.38 (above B+). Over the same period, the fraction of grades in the A range increased from 29.9% to 47.9%. Indeed, it was recognition of this trend that resulted in the adoption of the current grading policy in 2004 (effective for the 2005 academic year).
- GPA peaked in the 2002 academic year and subsequently fell sharply through 2005, anticipating by two years the application of the current grading policy. While the drop in GPA in 2003 is influenced by a change in the way Pass/Fail grades were recorded (see the discussion of Figure 2 below), both the average GPA and the fraction of A-range grades declined further in 2004, a year before the present grading policy was adopted.
- GPA increased somewhat after 2006, by 0.05 GPA points. The fraction of A-range grades continued to decrease a bit for a few years after 2005, but between 2009 and 2013, the fraction increased by 3.3 percentage points. These recent increases may reflect a change in monitoring the current policy.

The timing of the decline in average GPA and the fraction of A-range grades suggests that the implementation of the current grading policy was only partially responsible for arresting the increase in grades that had occurred over the previous quarter century. However, as outlined above, the timing does suggest a potentially important alternative understanding that helped guide us in formulating our recommendations. The decline in average GPA and the fraction of A-range grades began at about the time a vigorous debate started among the faculty regarding grading, grade inflation, and a potential change in grading policy. This discussion was precipitated by interest from the dean of the college in the issue, and ultimately resulted in the 2004 grading policy with numerical targets for the fraction of A-range grades. What the

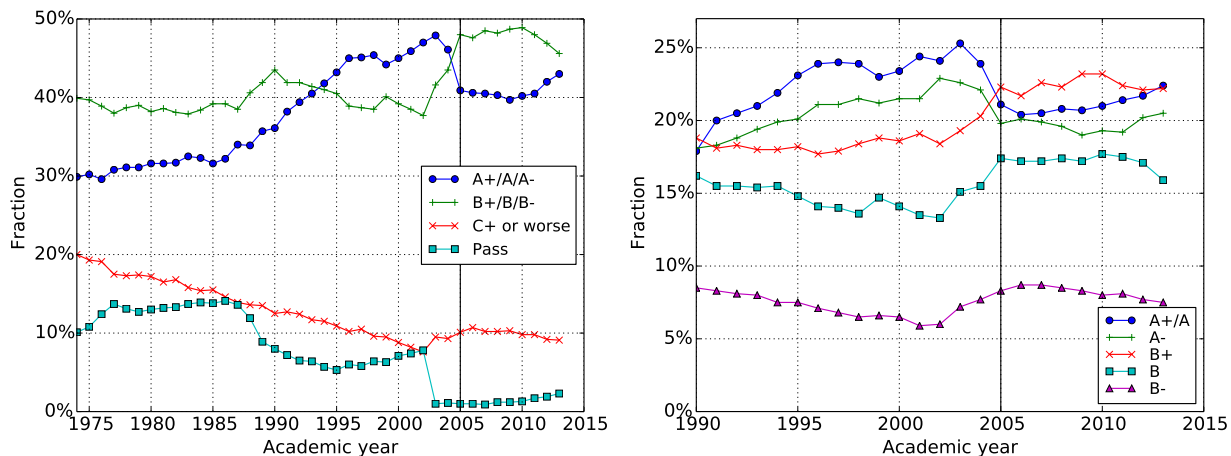


Figure 2: Fraction of A, B, and C-or-worse grades by year, 1974–2013

time pattern of results suggests is that much of the change resulted from the attention and discussion rather than from the hard guidelines themselves.

Clearly, guidelines can help keep the faculty focused on the importance of giving informative feedback to students through grades. However, we believe that the same result can be achieved through regular campus discussion at the departmental level. The dean of the college can evaluate on an ongoing basis whether departments are deviating from their past norms and suggest to such departments that they should examine why this might be the case in order to ensure that grades continue to play the role they are designed to have in providing feedback to students.

Grade Compression

Fine distinctions among students cannot be made with a very narrow range of grades awarded, so one potential benefit of the 2004 grading policy would be that a reduction in the fraction of A-range grades would increase the effective range of grades awarded. The left panel of Figure 2 contains plots over time of the fraction of grades that are A-range, B-range, and C-or-worse range, as well as the fraction of grades recorded as Pass.

We now mention a subtle point that complicates the interpretation of the data just prior to the implementation of the 2004 grading policy. Note in Figure 2 that the number of Pass grades drops sharply in 2003, from 7.8% to 1%. This sharp drop reflects a change in the policy with regard to how these grades were recorded. Prior to 2003, if a student elected to take a course Pass/Fail, a passing grade was recorded by the registrar simply as a Pass. Starting in 2003, the grade recorded by the registrar was the letter grade assigned by the course instructor; these grades were then covered on the students' transcripts, where they appeared as a Pass. In 2003 and subsequent years, the only grades recorded as Pass are those for which the course is Pass/Fail only, in which case the instructor does not assign a letter grade. The net result is that more B and C-range grades were recorded in 2003 and thereafter; this in turn results in a lower average GPA, since Pass grades are not included in GPA computation. However, it may have little effect on the fraction of A-range grades, since, given the ability of students to elect P/D/F status fairly late in the term, it is likely that only a small fraction of grades previously recorded as Pass grades would have been in the A range.

The time-series for the A-range grades shown in Figure 2 reproduces the raw plot in the right panel of Figure 1, and shows the steady increase in the fraction of A-range grades through 2003. This is mirrored by an equally steady decrease in the fraction of C-or-worse grades over the same period (from 20% of grades recorded—and presumably more, once Pass grades are taken into account—to less than 10% in 2003). Interestingly, the fraction of B-range grades was constant through the early 1990s and then declined by about 5 percentage points through 2002 as the fraction of A-range grades grew more rapidly.

Three changes subsequent to 2003 are clear:

1. The fraction of A-range grades declined significantly (by 7%) by 2005.
2. The fraction of B-range grades increased significantly (by 6.4%) by 2005.
3. The fraction of grades C or worse held steady (did not continue its trend decline).

We may attribute these changes to the discussion surrounding grading at Princeton that led to the 2004 grading policy, but it does not seem that they did much to improve the ability to make fine distinctions in student performance. A-range and B-range grades have accounted for around 90% of grades at Princeton since 2003.

Neither the shift from A-range to B-range grades nor the lack of change in the fraction of C-or-worse grades between 2003 and 2005 is surprising given the focus of the discussion (and the resulting policy) on the fraction of A-range grades. A more nuanced look at the data, contained in the right panel of Figure 2, shows the time series from 1992 of the grade distribution for A+/A, A-, B+, B, and B-. Usage of each grade in the A range decreased between 2003 and 2005 while usage of each grade in the B range increased over the same period. Thus, the faculty did not simply move some grades from A- to B+ to reduce the number of A-range grades awarded: the entire distribution of grades in the A-range and B-range moved down, as is apparent in the right panel of Figure 2.

Once again, the timing of these changes does not coincide with the implementation of the 2004 grading policy. But it does coincide with the beginning of discussion among the faculty regarding grading, again suggesting that encouraging continuing discussion among the faculty of grading and its use to provide meaningful feedback to students is likely to be a fruitful avenue going forward to encourage appropriate grading practices.

Variation across departments

The question of whether the grading policy has led to less variation in the grade distribution across departments is a more subtle one. As discussed in Section 2.1, we feel there is a difference between “consistent standards” described in the policy and “consistent grades,” but since the former is difficult to quantify across departments in a wide variety of disciplines, in this section we will examine the extent to which the grading policy has been successful in making *grades* more consistent across departments. This question is, of course, examined regularly by the Committee on Grading, but our analysis might provide a different perspective.

Overall, there has been some reduction since the 2002-2004 period in the disparity of grades across departments, but there is still significant variation. Figure 3 shows one view of the data, in which we show the fraction of A-range grades in each department, averaged over a three-year period ending in Academic Year 2013, compared with the same quantity for the three-year period ending in Academic Year 2004, just before the policy was implemented. The horizontal line at 35% indicates the target number established by the policy, and the 45-degree

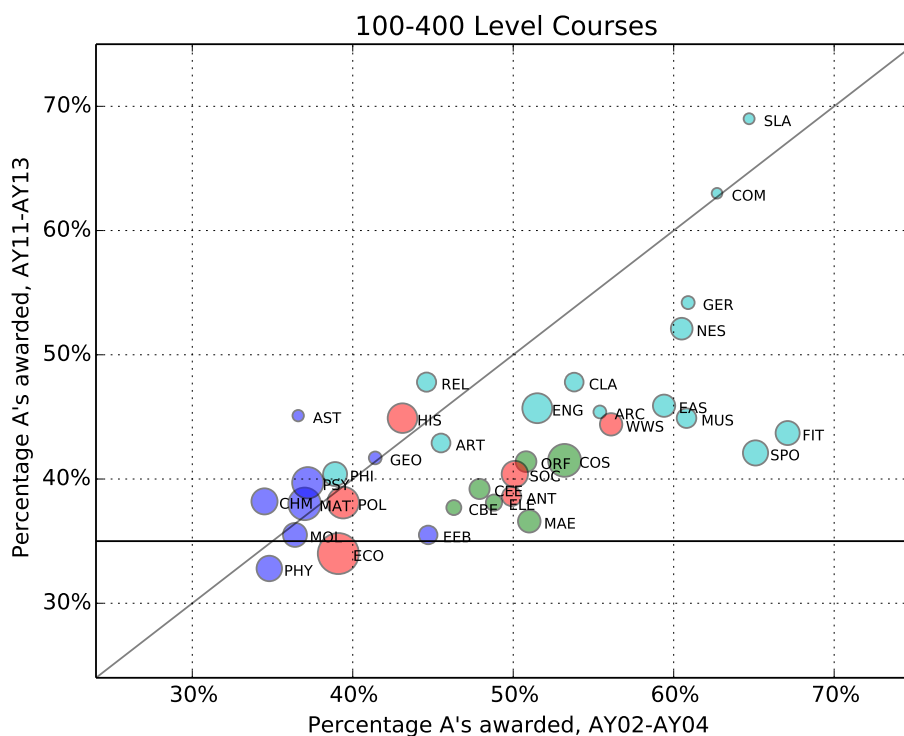


Figure 3: Percentage of A-range grades by department, for all graded work except junior papers and senior theses, shown as 3-year averages for Academic Years 2011–13, versus Academic Years 2002–4. The area of each symbol is proportional to the total enrollment in the AY11–13 period, and the departments are color-coded by division.

line indicates “no change” over the period. The area of the symbols in the plot is proportional to the number of grades awarded in each department. The figure shows that most departments have shifted closer to the 35% target, with the notable exception of two smaller departments in the humanities.

However, looking at the average across all grades given by a department (as prescribed in the 2004 grading policy) obscures some subtleties, in particular the sometimes large discrepancy between grades in introductory and upper-level courses. These discrepancies are illustrated in Figure 4, in which the grades are broken down between 100–200 level courses and 300–400 level courses. It is clear that the grades in upper-level courses are higher than those in lower-level courses. For lower-level courses, many departments average below 35% A grades, while for upper-level courses, no departments average below 35%. The fact that grades are higher in upper-level courses is not surprising. Students taking upper-level courses are largely self-selecting: they are often departmental concentrators who have discovered a particular interest in or aptitude for the subject, while students in lower-level courses are often exploring subjects for which many of them will find they have less interest or aptitude. This phenomenon has significant implications for the effect of the current grading policy (based on the departmental-averaged grade distributions) for departments with large service courses, compared with departments with fewer such courses.

Consider the following example. Mathematics has the largest enrollment in 100–200 level

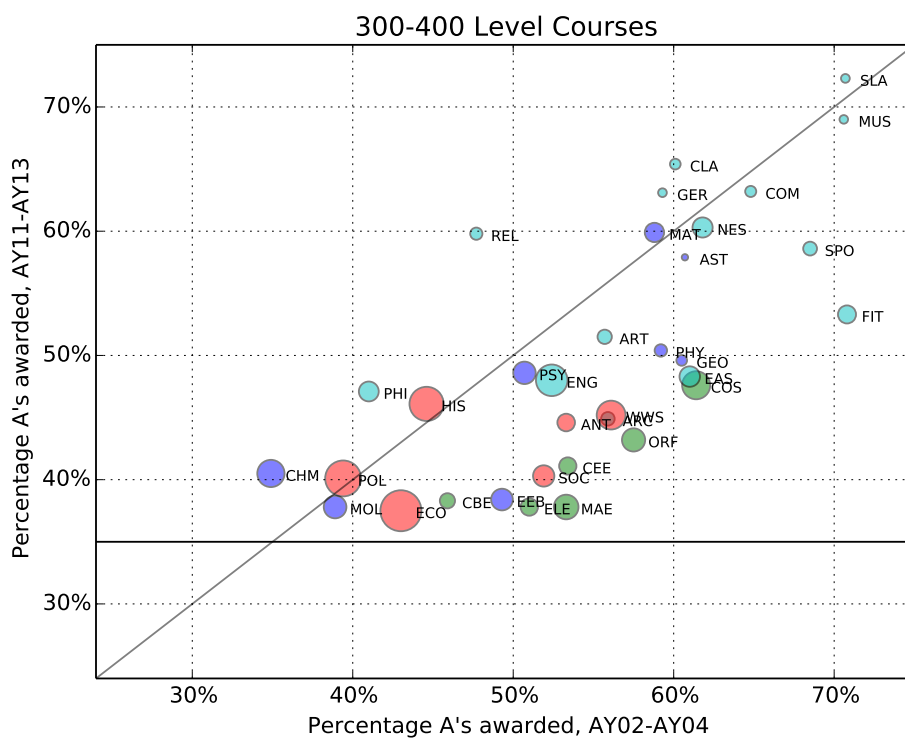
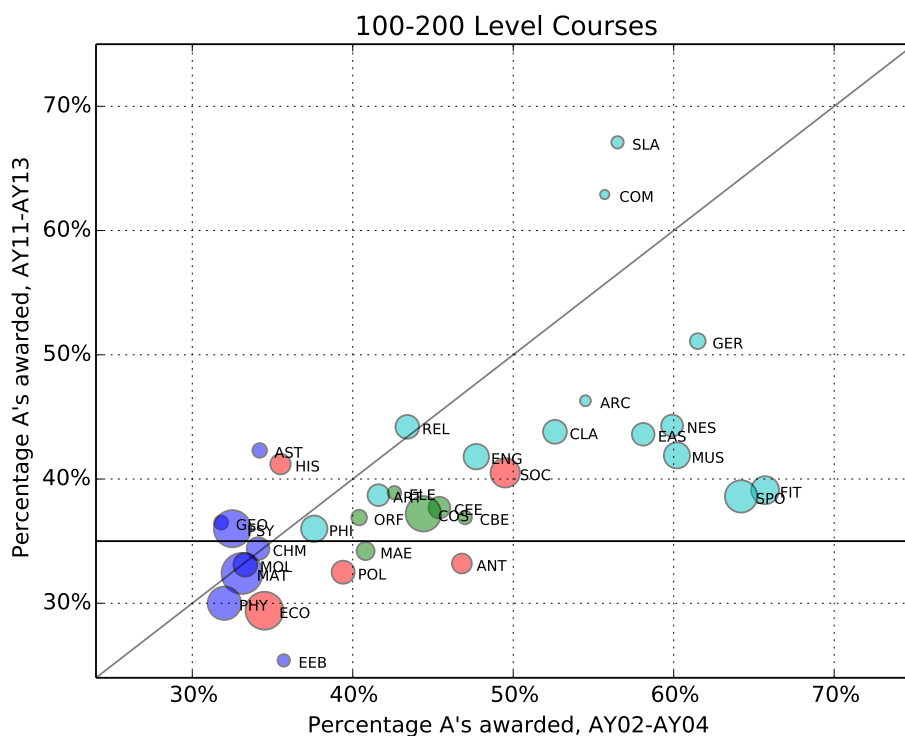


Figure 4: Percentage of A-range grades by department, shown as 3-year averages as in Figure 3, separated into 100–200 level courses (top plot), and 300–400 level courses, excluding junior papers and senior theses (bottom plot). The area is again proportional to the enrollment, and the departments are color-coded by division.

courses of any department, and the average fraction of A-range grades in these courses is 32.4% (averaged over the 3-year period ending in Academic Year 2013). In contrast, the fraction of A-range grades in *upper-level* Mathematics courses is 59.9%. Across the whole Mathematics department, the fraction of A-range grades is 38.0%. Looking only at the department-wide average, one might conclude that students concentrating in Mathematics should receive, on average, roughly the same fraction of A-range grades as those concentrating in Engineering departments such as CBE (37.7%), ELE (38.1%), or MAE (36.6%). However, this is grossly misleading, since the *upper-level* courses in the Engineering departments have a much lower fraction of A-range grades than those in Mathematics and Physics: 38.3% for CBE, 37.8% for ELE, and 37.8% for MAE, compared with 59.9% for Mathematics and 50.4% for Physics. Ironically, many of the students taking the introductory Mathematics and Physics courses are in fact Engineering students, so these students get hit twice: first in the introductory Math and Physics courses, and then in their own departments, which are “limited” in the number of A-range grades they can give because Engineering departments do not offer service courses with large enrollments.

Note that the grading policy has had little effect on the grades in the Mathematics department (for both lower-level and upper-level courses, the points in Figures 3 and 4 are on the 45-degree line). However, it has had a substantial impact on the Engineering departments, which have significantly lowered the fraction of A-grades in both lower-level and upper-level courses. Thus, while it appears from the overall averages in Figure 3 that the grading policy has reduced the disparity across departments, closer examination of Figure 4 reveals that, in some cases, it has made the disparity worse.

Thus, while the grading policy has succeeded in making the overall fraction of A-range grades more uniform across departments, in our view this is not a reliable measure of fairness across departments. Because some departments by their very nature offer service courses with large enrollments, while others do not, comparing department-wide averages is not even a reliable measure of “consistent grades” across departments, much less the more subtle and elusive notion of “consistent standards” across departments.

On balance, we feel that emphasis on moving various departments further toward a common grade distribution is misplaced. There is a long history of different departments using different distributions, and we understand that this is not a Princeton-specific phenomenon. Our educated guess is that the rank ordering of grading by departments does not differ much across our peer institutions. This is well understood by the world at large (e.g., graduate schools, employers), and it does not appear to cause problems, at least none that surfaced in our discussions with a wide range of groups. Our view is that grades within departments need to be meaningful in providing accurate feedback to students but that this does not require identical grade distributions across departments.

3 Public perception of the current policy

3.1 Survey results

The committee commissioned three surveys: for current students, current faculty, and the public at large. The surveys were administered and data collected by Vice Provost Jed Marsh during February and March 2014.

The full survey results are included in Appendix C. To summarize briefly, among the 45% of undergraduates who responded to the survey, a majority indicated support for the goals

of the grading policy, but expressed strongly negative views on the policy's implementation, effectiveness, and influence on their academic experience at Princeton.

Faculty sentiment, with about half of the faculty responding, was roughly equal on both sides of the question, though with significant differences among divisions: the humanists were more strongly opposed to the policy than were the natural scientists. Public sentiment (primarily alumni and/or parents) was strongly in favor of the goals of the grading policy, but there was an almost equally strong belief that it has been ineffective.

In addition to the numerical responses, there were a large number of free-text responses. The student survey included three free-response questions, discussed briefly below:

Have you personally experienced any negative effects from the grading policy?

1429 text responses, with 87,883 words (average 61, longest 5,037). The word "negative" occurred 211 times; the word "positive" occurred 24 times, but in context many apparent 'positives' were not. For instance: "I do not think that the grading policy is a positive contribution to the school in any way." "There are zero positive effects."

Have you personally experienced any positive effects from the grading policy?

1192 responses, 25,059 words (average 21, longest 1778). Again, "negative" occurred 36 times, "positive" occurred 124 times, but in context, many apparent 'positives' were not: "I have not experienced the positive effects of the grading policy." "No I have experienced no positive effects."

Do you have specific suggestions for improving Princeton's grading policy?

1274 responses, 70,696 words (average 55, longest 5412). A list of "collocations," that is, word pairs that occur unusually frequently, is suggestive:

grade deflation; grading policy; Get rid; peer institutions; graduate schools; Ivy League; Princeton students; grade inflation; quality work; across departments; grading policies; graduate school; grading system; feel like; deflation policy; playing field; current policy; grad school

Several themes emerged from the student responses, and we provide a few sample responses below. Some of the comments revealed poor behavior on the part of the faculty:

I received a 91 on a midterm exam in a [particular department] course this past fall (my concentration), but the 91 was scratched out and replaced with an 88. When I asked my professor why he reduced my score, he told me that normally the paper would be an A-, but due to grade deflation, he was forced to lower several students' grades to a B+.

The grading policy is particularly unreasonable in introductory language courses. On the first day of classes, my [language] teacher said that only 3 of us in a class of 11 would receive As. This often means that despite receiving an overall grade of 90+ a student cannot receive an A-grade because some other student got a 91 or 92.

Another common theme was that the grading policy harms the spirit of collaboration:

I have experience[d] multiple negative effects from the grading policy. Because of grade deflation it has been extremely hard to find any kind of collaborative environment in any department and class I have taken at Princeton. Often even good friends of mine would refuse to explain simple concepts that I might have not understood in class for fear that I would do better than them. I have also heard from others about students actively sabotaging other student's grades by giving them the wrong notes or telling them wrong information. Classes here often feel like shark tanks. If I had known about this I very probably would have not attended Princeton despite it being a wonderful university otherwise.

Some students described how their low grades have had serious consequences, which they blame (rightly or wrongly) on the grading policy:

I had to drop being Pre-Med here because the grades I was getting in the sciences were too low. I was getting low grades not because I didn't understand the material, but because the curve was getting messed up by kids who were very advanced in chemistry and taking Intro to Chem and getting 100's on the exams. Now my parents have to help me pay for a post-bac program so that I can take the sciences elsewhere post-graduation because Princeton didn't allow me to take the necessary next step to realizing my dream by giving me unfair grades in the sciences. My sister went to [a peer institution] and was pre-med there, and even though I have been consistently better than her grade wise growing up, she was able to receive A's in all the science courses and attend [a top medical school]. I don't even know if I have a shot at [that medical school] because of Princeton. Not because I'm not qualified or a good applicant, but because Princeton's grading policy makes me look like a poor applicant compared to other students applying with incredibly high GPA's.

I earned a college scholarship at the end of high school. To keep the scholarship, I had to maintain a 3.4 GPA throughout college. I did not have a 3.4 my first semester due to grade deflation in large introduction classes and lost my scholarship. / Also, many internships have 3.5 or 3.6 GPA cutoffs. They don't care what school you go to or that Princeton has grade deflation. Your application isn't considered if you don't make the cutoff.

Other students have had positive experiences with the grading policy. One common theme that emerged was that, because of the policy, students appreciate A grades more:

On the whole I've gotten a relatively high GPA and Princeton's grading policy has made those grades count more, both in terms of applying for future opportunities and for my self-esteem.

I do believe that most of my grades have been fair, and the grading policy helps to know they are not inflated.

Another theme was that the policy provides an incentive for students to work harder:

I think the rigorous standards (either natural to Princeton or caused by the grade deflation policy) have definitely pushed me harder to do better work, especially in the classes I enjoy. I think one of the unmentioned effects is the clarification I feel because of grade deflation. In some classes, I'll feel very stressed trying to produce quality work, but in others, I enjoy putting the extra time/effort in and want to be challenged. Unsurprisingly, this helped me figure out which departments and disciplines I enjoy, and which I would rather avoid.

One student's honest response summed up the situation nicely:

The grading policy provides students (myself included) with an easy excuse anytime we receive grades that we don't feel reflect our work.

3.2 Views from the USG

The committee met with members of the Undergraduate Student Government (USG), who expressed four major concerns with the grading policy:

1. It deters admitted students from matriculating.
2. It may cause students to select departments based upon perceived grading tendencies, and may harm concentrators in small departments or departments in which mostly upper-level courses are offered.
3. It has led to emphasis on the letter grade rather than substantive instructor feedback.
4. It disadvantages Princeton students in the external marketplace, where some employers have GPA cutoffs and are likely to treat Ivy League GPAs as providing similar information.

The USG also shared key insights from the Academic Life Total Assessment (ALTA) administered in January 2012. Among the 2567 undergraduate respondents, 46% said they had received justification for a grade based on the grading policy and 23% said they were not getting enough feedback. Students in both the survey and associated focus groups expressed interest in greater frequency of assessment and more informative feedback, especially earlier in the term. The survey also raised concern over the psychological impact of grades, to be addressed in Section 4.1.

4 Side effects of the current policy

4.1 Anxiety among the students

One of the negative side effects of the grading policy has been its contribution—in perception at least—to the anxiety about grades and indeed about themselves that many students experience while at Princeton. Chemistry professor Michael Hecht expressed the idea well by describing the campus as a pressure cooker and suggesting that the grading policy has kept the

lid on tight. “I believe that student well-being, enthusiasm, and peace of mind are important—both for student life in general and for academic success,” he said, reflecting on his four years as Master of Forbes College. “I would rather our students be motivated by love of learning than anxiety about grades. [The] current grading policy is detrimental to the learning experience and to the campus environment in general. Let’s fix that.”

Professor Hecht’s comments align with well-established research, which suggests that an optimal level of anxiety is essential for performing to the best of one’s ability on academic work, but that an elevated level of anxiety can block mental performance, lead to procrastination, and be associated with psychosomatic symptoms, alienation, and diminished self worth.

We should note that the committee received no hard evidence with which to substantiate absolutely that student anxiety has increased as a direct result of the grading policy during the years in which the policy has been in effect. Although some data about mental health assessments might be available, the committee considered it unlikely that trends could be identified that would separate grading policy effects from other factors associated with economic conditions, family backgrounds, and job markets.

However, the committee did receive considerable evidence to support the argument, unsurprisingly, that grades, anxiety about grades, and students’ sense of psychological well-being (or lack thereof) are closely related. For example, the ALTA survey showed that 40% of students said grades have a “strong psychological impact on life outside of classes” while another 48% said grades have some impact.

From the various faculty, staff, and students who met with the committee, we identified two possible ways in which the grading policy may be exacerbating student anxiety. One is students’ perception that only 35% of those enrolled in a particular class can receive A-range grades, no matter how hard everyone works. Such a perception would undoubtedly contribute to the “culture of competition” that is already so prominent on campus, aggravating it by reinforcing the idea that students are locked in a zero sum game and should view their classmates as competitors rather than as friends and colleagues.

The other possibility is anxiety driven by the belief that a 35% cap on A grades adds to, rather than reduces, whatever arbitrariness may be present in the relationship between effort and grades. If effort that truly leads to top-quality work does not result in equivalent grades, the psychological impact can be a kind of fatalism or at least a heightened level of uncertainty about how well one is doing and whether the reward system is functioning rationally.

4.2 Competitiveness outside Princeton

In recent months, two of our principal institutional competitors, Harvard and Yale, have been in the national news for the high percentage of A-range grades that they award. The press and general public opinion have not been kind to the revelations of such lax standards, but one might argue that Harvard and Yale have the last laugh since there is no evidence—or at least we have not been presented with any—that students at these universities are adversely affected by transcripts peppered with A’s. All educational institutions are different, sometimes in subtle ways, and we can be sure that comparisons between Princeton and such other Ivy League institutions are between apples and pears, if not apples and oranges. Nevertheless, it is inevitable that people—applicants, current students, parents, admissions committees at graduate and professional schools, fellowship-awarding bodies, and employers—will compare our current policy with the public announcement of grading practices elsewhere. For example, Yale reported in February 2013 that by the spring of 2012, 62% of undergraduate grades were A’s and A–’s, while the dean of undergraduate education at Harvard, basing his remarks

on data from the fall of 2012 and several previous semesters, reported at a faculty meeting in February 2013 that “The median grade in Harvard College is indeed an A–. The most frequently awarded grade in Harvard College is actually a straight A.” Whether anything will change at Harvard and Yale remains to be seen.

Does the discrepancy between our policy and the reality in Cambridge, New Haven, and perhaps elsewhere make us—and Boston University and Wellesley College, two institutions that have also attempted to do something about rampant grade inflation—less competitive? Other sections of this report consider the effect the policy may have on applicants to Princeton (Section 4.3) and the stress to which it may subject enrolled students (and, secondarily, their families) during the school year (Section 4.1). Here we consider how it may affect our undergraduates’ success in gaining admission to graduate and professional school, winning major external fellowships, being placed in plum ROTC assignments, and landing jobs.

As far as graduate school is concerned, it is not evident that Princeton’s grading policy has any effect. Students typically attend graduate school in subjects for which they have particular aptitude and in which they have done particularly well as undergraduates. While departments sometimes make first cuts in their applicant pool based on such factors as GPA, we have no reason to believe that Princeton students are failing to gain admission to Ph.D. programs. Of particular interest is the related matter of medical school, since would-be doctors evince an unusually high degree of anxiety about getting in, especially to the most competitive programs (for instance, see one student’s comment on page 12). In a detailed report to the committee, Kate Fukawa-Connelly, Director of Health Professions Advising, presented evidence that our grading policy does not measurably affect our students’ frankly remarkable success; if the policy were not in effect, the students who blame it for not achieving the results for which they hoped—being accepted to top medical schools is necessarily a very difficult proposition under any circumstances—would presumably find something else on which to pin their dissatisfaction.

Another likewise difficult proposition is winning a major postgraduate fellowship, of which the most prominent ones that take into account the whole applicant and his/her overall academic record, are the Rhodes, Marshall, and Gates Cambridge Scholarships. It is true that it is difficult to get an interview for a Marshall Scholarship in particular without having a GPA of 3.9 at the time of application. It is, however, also true that the number of available scholarships—like the number of available places at Stanford Medical School—is severely limited. While Deirdre Moloney, Director of Fellowship Advising, sees a positive correlation between Harvard’s high grades and Harvard’s typically (though not inevitably) more robust number of recipients of major external fellowships, it is not obvious to the committee that removing the grading policy would in itself increase the number of winners Princeton has each year; possibly it would increase the number and range of Princetonians whose viability fellowship-granting bodies might seriously debate, but it is not likely to significantly change how many ultimately prove successful, a number that will in any event be low. The notion that one or two more of our very top students might in a given year win a major external fellowship is not a good reason to change this policy. The fact remains that, as with admission to medical school, Princeton’s record in the Rhodes, Marshall, and Gates competitions is impressive, and it is worth noting that the American secretary of the Rhodes Trust has explicitly stated that the grading policy “is not harming Princeton’s students” and that “We hope very much Princeton retains its standards, and frankly, would like to see other universities follow its lead.”

The relationship between GPA and employment—both in the summers during college and after graduation—is more complicated. On the one hand, the committee was surprised to learn that many of the most competitive jobs (e.g., in finance) require candidates to report

their GPA but without submitting a transcript. Under such circumstances—and especially for those positions that have firm GPA cutoffs, which are in fact sometimes mandated by the Department of Labor—it is clear that more candidates will be “in the running” from Harvard and Yale than from Princeton. However, here, too, there are only so many jobs on offer at the top firms, so while it is possible that a few different Princetonians would get jobs at, say, Goldman Sachs if grades were higher, the committee heard evidence that the actual number of Princetonians in such jobs would be the same. Indeed, although one prominent banker issued strong words about how the policy hurts Princetonians, close examination did not persuade us that this is so. In addition, if one looks at the overall picture and does not concentrate solely on the students at the top of the class, Princetonians appear not to have unusual difficulty convincing potential employers to hire them for jobs at companies that are a notch below the most elite.

It should be noted that there is one constituency—a very small one at Princeton—for which there is a undeniable link between GPA and postgraduate position and, furthermore, for which this position is neither optional (as with a fellowship) nor more or less easily convertible (as with a job): ROTC. Cadets and midshipmen across the country are ranked lineally, and an officer’s number, for which GPA accounts for approximately two thirds of the calculation, directly affects his/her first assignment and, at a larger level, entire career. While it would be unreasonable for Princeton to change its grading policy as a result of a choice made by only a small number of students in each graduating class, ROTC comprises a special group of students whose issues deserve to be taken seriously.

In short, with the small exception of ROTC, the best reasons to change Princeton’s grading policy have more to do with psychological factors and campus atmosphere than with any tangible negative effects it has on the prospects of our students outside “the Orange Bubble.”

4.3 Admissions yield

The perception that the number of A-range grades is limited sends the message that students will not be properly rewarded for their work. During the application process, students and parents consider the possible ramifications in terms of reduced future placement and employment potential. This concern is acute among admitted students who are making their final college choice, including athletes. Coaches find the perception of the grading policy a significant obstacle to recruitment, making it more difficult for them to attract the best student-athletes to Princeton.

Janet Rapelye, Dean of Admission, reports that the grading policy is the most discussed topic at Princeton Preview and explains that prospective students and their parents see the numerical targets as inflexible. The committee was surprised to learn that students at other schools (e.g., Harvard, Stanford, and Yale) use our grading policy to recruit against us.

In Dean Rapelye’s view, removing the numerical targets would go a long way towards alleviating these concerns and would make the grading policy less of an issue among prospective students and their families. People fixate on numbers, and the very existence of a numerical guideline such as 35% serves as a lightning rod, giving (perhaps wrongly) the impression of inflexibility. Removing the numerical target without changing the intent of the policy would solve many of these issues.

4.4 Impact on freshmen

The committee explored the possibility that the grading policy may be contributing to the significant pressure placed on freshmen as they move from high school to Princeton.

The transition from high school to Princeton is challenging on many levels. Adjustment disorder continues to rank as a top diagnosis for students seeking Princeton Counseling and Psychological Services. According to John Kolligian, Director of University Health Services, National College Health Assessment data reveal that Princeton undergraduates experience higher levels of stress than the undergraduate reference group, with around 47% of Princeton students reporting “more than normal” levels of stress and around 10% reporting tremendous stress. The data also reveal that stress is the most significant factor negatively influencing academic performance.

With these issues in mind, the committee considered whether the grading policy was contributing to stress in the freshman year as students from diverse backgrounds transition into the rigorous academic programs at Princeton. Arguably, part of the “culture shock” in coming to Princeton is that the average GPA for freshmen is 3.24, a value significantly lower than what most students earned in high school. The committee investigated possible solutions including pass/fail grades or covering grades in the first semester or year. These options will be discussed in Section 5.4.

5 Alternative policies considered

Here, we mention several possible changes to the current policy that we considered but ultimately did not end up recommending.

5.1 Change the currency

Until 1969, Princeton used a 7-point numerical grading scale rather than the letters A to F that all current students and nearly all members of the faculty take for granted and that most, though not all, of our peer institutions employ as well. Older alumni can be attached to the 7-point scale, and the news that the ad hoc committee had been formed caused some of them to compose letters to suggest that Princeton return to it. Perhaps the most thoughtful commentary came from the pen of Richard Etlin '69 *72 *78, a professor emeritus at the University of Maryland, who wrote to the committee in November 2013 and whose piece was published in slightly shortened form in the *Princeton Alumni Weekly* in January.

The committee took the issue seriously. In the end, however, we are not convinced that changing the currency, as it were, would affect the distribution of grades—and even if it did, slightly, the resulting confusion would not be worth it. It is true that normalization is a trivial mechanical operation—converting between a 7-point scale and a 4-point one, which is in effect what A-F is, is hardly difficult—but we are not comfortable with the assumption that all, or even most, readers of applications would understand and look kindly on a non-standard system and take the time to make the adjustment. (Our closest peer with a non-standard system is MIT, which has a 5-point scale; the registrar’s page provides a brief section titled “Conversion to a 4.0 Scale,” the gist of which is that 1 gets subtracted from every grade that is not an F.) Furthermore, taking the current policy, which has had the unfortunate effect of making Princeton look anomalous rather than visionary, and replacing it in whole or in part with a different anomaly would be unlikely to help matters—and might, furthermore, continue to

suggest to the outside world that Princeton cares more about its grading policy as such than about articulating something truly important, namely what exactly the differences are between a piece of A work and one that earns an A–, not to mention a B+ or a B or a C or a D.

The University grading policy draws a sharp line between A-range and non-A-range grades. This has had the effect of making students even more worried than usual about the (necessarily arbitrary) distinction between an A– and a B+ while also giving faculty little incentive to make use of the entire grading scale and give grades below, or much below, a B+. The committee has no wish to increase the number of C's and D's for the sake of show but hopes that, whatever policy is adopted for the years ahead, faculty will feel comfortable making distinctions in quality other than between an A and an A– and between an A– and a B+.

There is, to be sure, also the matter of the A+, which is supposed to be the mark of truly exceptional work. Questions abound. Should the A+ be more widely employed or should some other new “super grade” be added to the grading scale? Should faculty really be required to complete “A+ Statements”? And should an A+ count again as a 4.3, as it did until 2000? Members of the committee do not all agree on the answers, though on the whole we are satisfied with the current state of affairs.

There is one last thing to note. Princeton is unusual in calculating GPA as a purely internal matter: the numbers are not made public, either on transcripts or elsewhere. Students thus calculate their own GPA and report it as they wish. It is beyond the charge of the committee to pronounce on the wisdom of this *laissez-faire* attitude, which might seem at odds with the fierceness with which the percentage of A-range grades has sometimes been policed; this—like the matter of A+'s—is an issue the University may wish to revisit.

5.2 Intentionally inflate grades

One possibility that the committee considered was to actually encourage grade inflation, so that the transcripts of Princeton's students would be more in line with those of our peer institutions. Indeed, inflating grades could have many positive side effects, as discussed in Section 4: it could relieve anxiety among students, make them more competitive for external fellowships, give them better job placement after graduation, etc. Some students and alumni expressed the view that, by giving lower grades than our peer institutions, we are actively harming our students.

However, as discussed in Section 4, the committee saw no clear evidence that our students are actually being harmed by having lower GPAs than their peers at Harvard and Yale (with the possible exception of a small number of ROTC students). Furthermore, we felt the deleterious effects of grade inflation are more insidious and serious than these more superficial advantages.

The committee acknowledges that removing the numerical targets may lead to some upward creep in grades. This is why our recommendations include continued monitoring and reporting of grades by the dean of the college, as discussed in Sections 1.2 and 2.2.

5.3 Report more information on the transcript

The committee spent a lot of time discussing what an A grade means, and possible alternatives or supplements to letter grades. In spite of the official definition of an A grade, there is no objective definition that is sufficiently operational that everyone would agree on its meaning and could apply it uniformly; indeed, the only thing that one can say for sure is that an A is better than a B.

The committee discussed a variety of alternatives that would either eliminate letter grades or make them more meaningful. Among the options:

1. Publish the percentages of A's in each class, or the percentages of all grades, perhaps in conjunction with other courses in the same department, division, or the whole University. These percentages are currently provided to individual faculty, so they can compare their grades with those of their colleagues; the information could be made more widely available—to students, on transcripts, etc.
2. Provide each student (and transcript) with the student's ordinal position within the class, either precisely or some suitable percentile. Princeton already groups students by quintiles, for better or worse; this information could be used in place of or to supplement letter grades.
3. Provide two GPA computations, inside and outside. The outside grade would be in some unspecified way normalized against our peers, notably Harvard and Yale. This would require sharing information with other institutions and would seem to be fraught with perils.

The committee did not feel that any of these options would alleviate anxiety around grades or soften the competitive atmosphere on campus.

5.4 Treat freshman year differently

Several peer institutions have used covered grading (MIT and Johns Hopkins) or pass/fail grading (Caltech and Swarthmore) in the freshman year. Each policy has subtle differences, but the basic idea is to report only pass/fail grades on the transcripts for the first semester of freshman year. Similarly, the earned grades do not count towards the student's overall GPA. "Covered grades" mean that the students themselves and authorized university personnel have access to the actual grades. At some institutions, the students may request that their actual grades be sent to potential employers. Interestingly, Johns Hopkins recently voted to repeal a covered grading policy because the policy created several significant academic problems. For example, in sequential courses, the students were often underprepared for the second semester, presumably because the first semester "did not count." Grades provide strong incentives to work hard, even to freshmen.

Overall, the committee foresaw several academic complications that might be created by pass/fail or covered grades and came to the conclusion that the freshman year adjustment problem, particularly for certain groups, is a complex issue that is not likely to be solved by simple changes to the grading policy. In the end, the members concluded that the issue was beyond the scope of the committee and that the freshman year transition may merit a more in-depth analysis.

A Committee members

Name	Title
Clancy Rowley (chair)	Professor of Mechanical and Aerospace Engineering
Henry Farber	Hughes-Rogers Professor of Economics and Director of the Industrial Relations Section
Devin Fore	Associate Professor of German
Alison Gammie	Senior Lecturer in Molecular Biology
William Gleason	Professor of English and Chair of the Department of English
Joshua Katz	Professor of Classics
Brian Kernighan	Professor of Computer Science
Bess Ward	William J. Sinclair Professor of Geosciences and Chair of the Department of Geosciences
Robert Wuthnow	Gerhard R. Andlinger '52 Professor of Social Sciences, Professor of Sociology, and Director of the Center for the Study of Religion
Elizabeth Colagiuri (staff)	Associate Dean of the College

B Charge to the committee

The evaluation of student work is a critical component of a liberal arts education. Faculty members expect students to submit their best work, and in turn, students expect faculty members to provide timely, thoughtful and fair feedback that will foster their intellectual growth.

Nearly ten years ago, in April 2004, the Faculty adopted a University grading policy. The policy recommends that each department, over time, award no more than 35% of A-range grades for course work, and no more than 55% of A-range grades for junior and senior independent work. The Faculty Committee on Grading has stated that

The grading policy has two fundamental objectives. The first is that the Princeton faculty wants grading to be done in such a way that students in one academic department can expect to be graded according to the same standards as students in any other academic department. . . . The second reason for the adoption of the grading policy is the conviction that students deserve clear signals from their teachers about the difference between their ordinarily good work and their very best work.

Since the implementation of the policy ten years ago, the number of A-range grades awarded across departments has become much more consistent. Likewise, the grade inflation of the late '90s and early 2000s has been halted. Yet concerns persist that the grading policy may have had unintended impacts upon the undergraduate academic experience that are not consistent with our broader educational goals.

The University periodically reviews major policies to determine whether they continue to serve its mission. The tenth anniversary of the grading policy's enactment is an appropriate time to undertake such a review. I have accordingly asked a committee of faculty members to spend the next year exploring a set of broad questions about the evaluation of student work:

1. Do the objectives of the Grading Policy, as recapitulated by the Faculty Committee on Grading, remain the appropriate ones against which to judge Princeton's assessment practices?
2. Does the current Policy achieve the University's pedagogical goals effectively and with as few negative consequences as possible, or are there better ways to do so?

I would ask the Committee, in the course of its deliberations, to consider any available evidence about the effects that the current policy might have on levels of feedback that students receive on academic work; the success of Princeton students with regard to employment processes and graduate school admissions; and the attitude of students toward their academic work and their peers. I would also ask the Committee to consider whether a grading policy might achieve Princeton's aims more effectively, and with fewer side effects, if it were to focus on clear standards, thoughtful rubrics and effective feedback, with less emphasis on numerical targets.

I expect that the Committee will consult broadly with the University community about the questions pending before it. I will ask the Committee to conclude its deliberations by recommending to the faculty any changes to the University's grading policies that would, in the Committee's judgment, improve our assessment practices, our pedagogy, and our general learning environment.

C Survey results

Survey Response Rates

Faculty

Total Respondents	467
Full Professor	219
Associate Professor	33
Assistant Professor	73
Senior Lecturer	21
Lecturer	105
Instructor	5
Other	11
Humanities	157
Social Sciences	109
Natural Sciences	111
Engineering	74
Other	16*

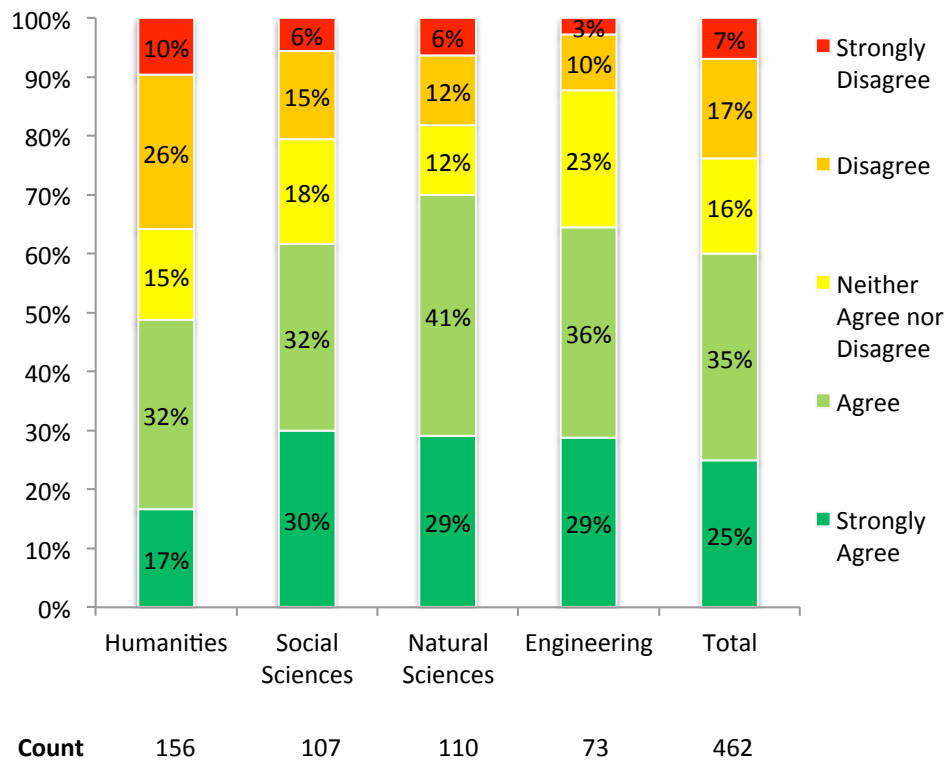
Students

Total Respondents	2423
Senior	554
Junior	545
Sophomore	623
Freshman	701
A.B.	1812
B.S.E.	611
Humanities	168
Social Sciences	397
Natural Sciences	279
Engineering	424
Undeclared	1155

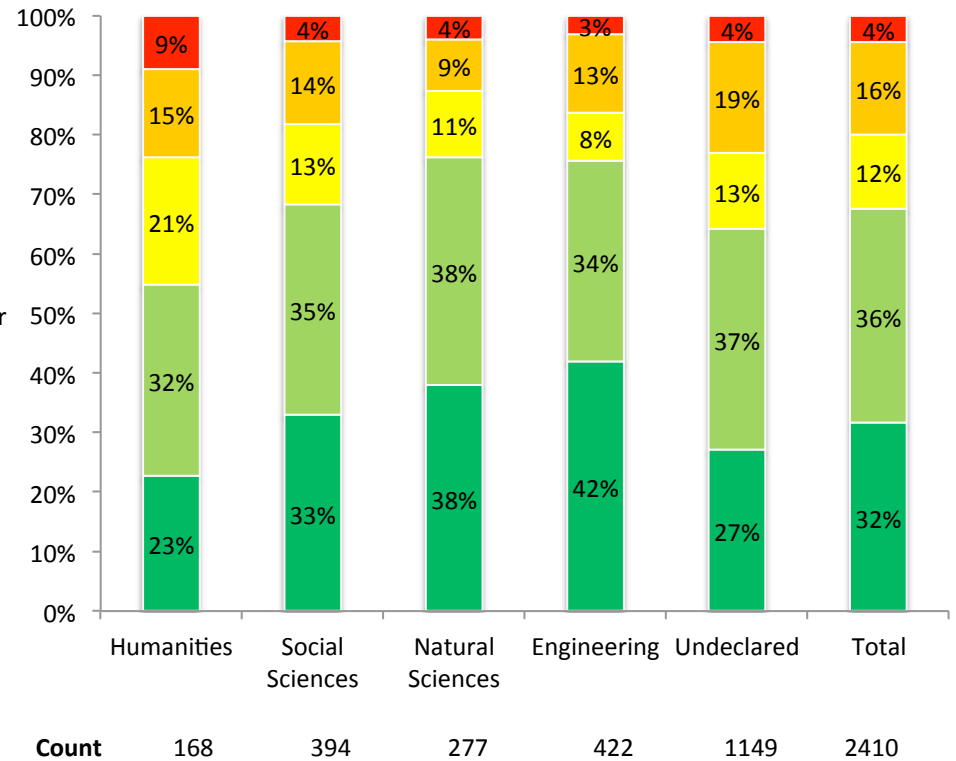
*Other responses have been included in totals, but not broken out in the graphs that follow

Q1A. To what extent do you agree or disagree with the following statement?
 Students in one academic department should be graded according to the same standards
 as students in any other department.

Faculty

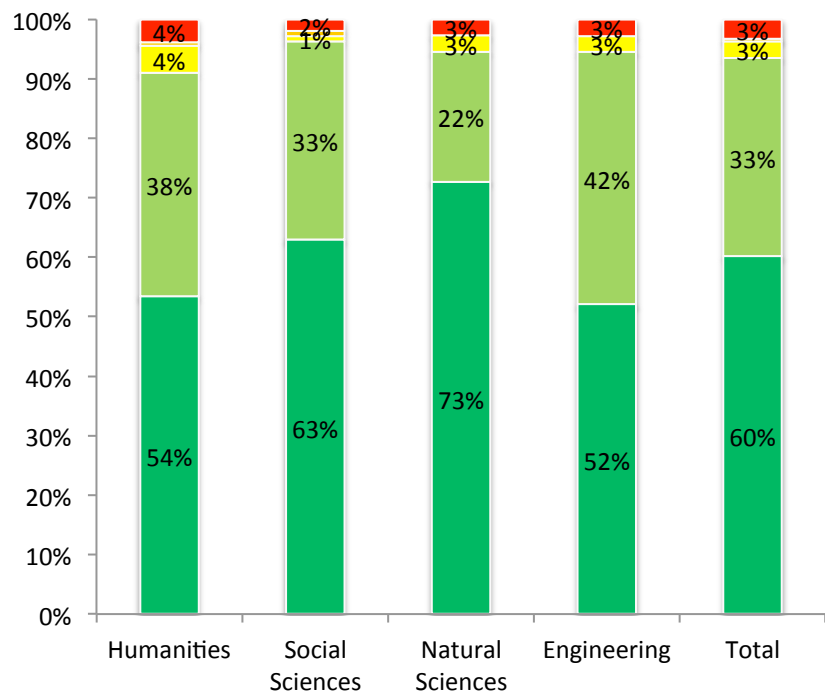


Students



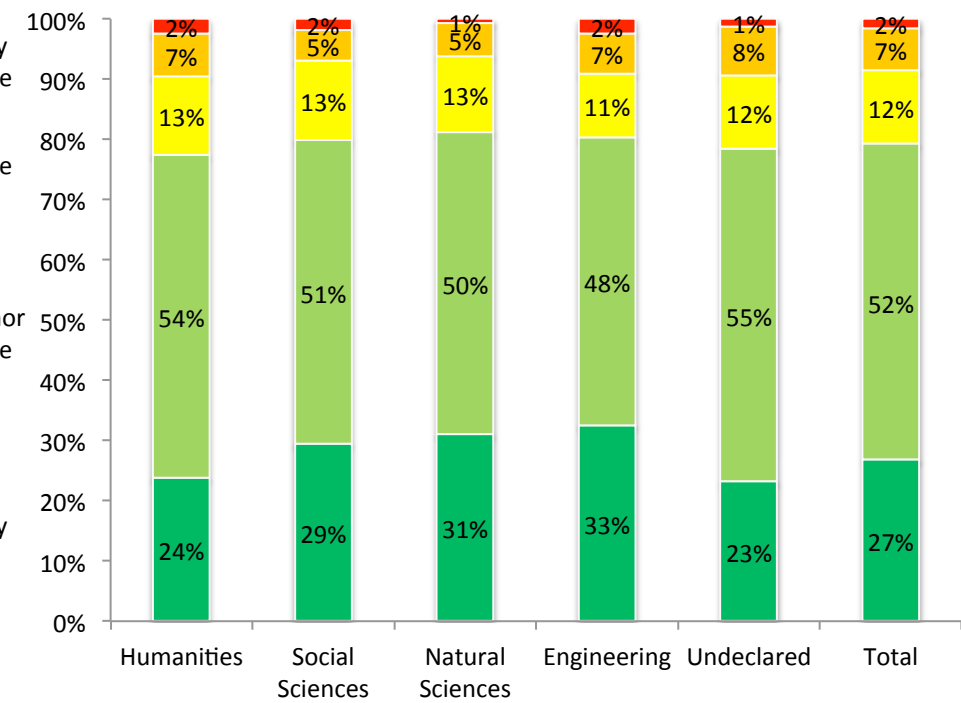
Q1B. To what extent do you agree or disagree with the following statement?
 Grades should give clear signals about the difference between ordinarily good work and the very best work.

Faculty



Count 157 108 110 73 464

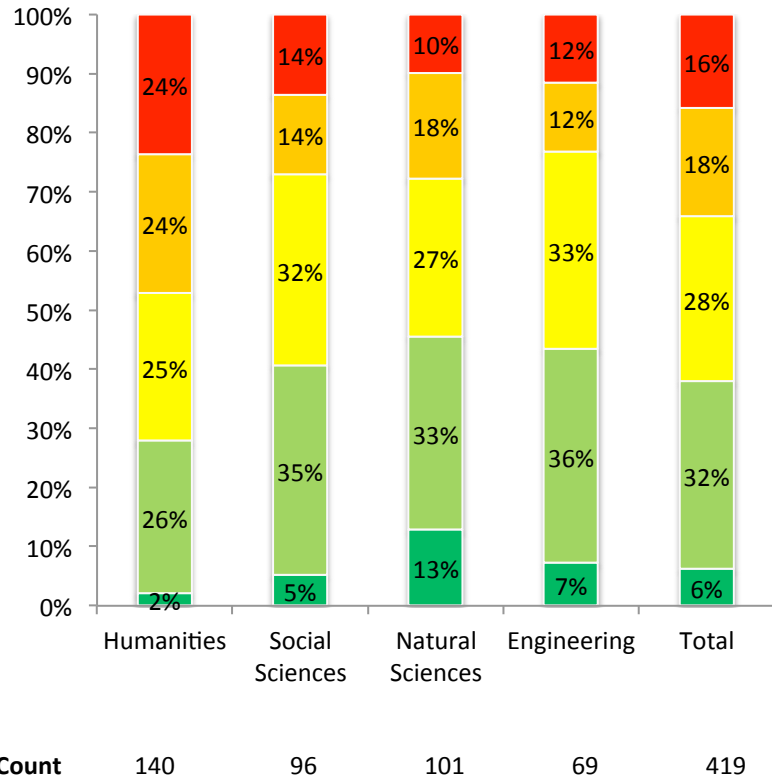
Students



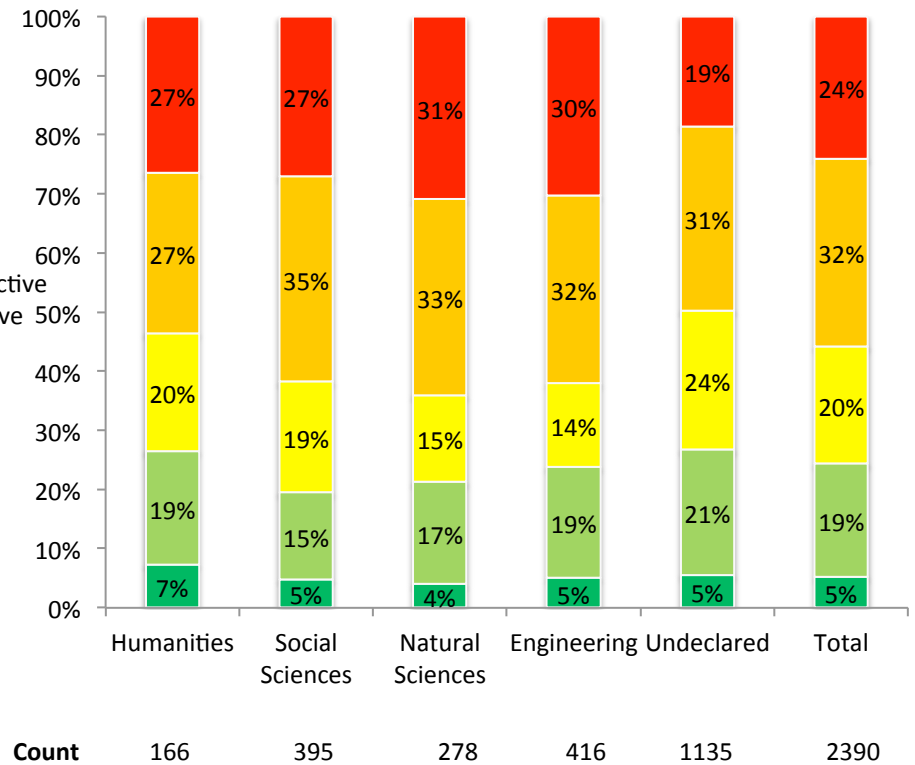
Count 168 394 277 421 1147 2407

Q2A. How effective do you feel the grading policy has been in meeting the following objective?
 Students in one academic department are graded according to the same standards
 as students in any other department.

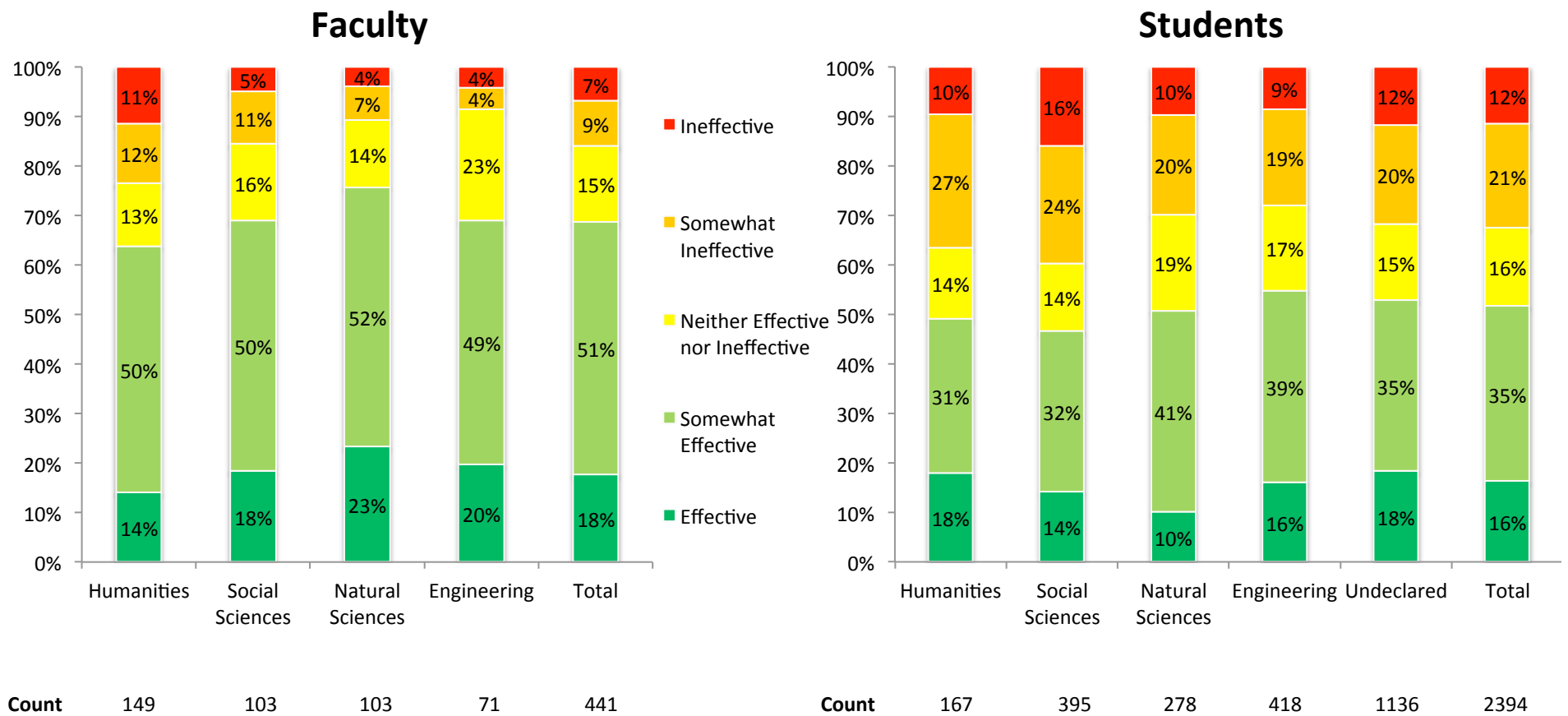
Faculty



Students

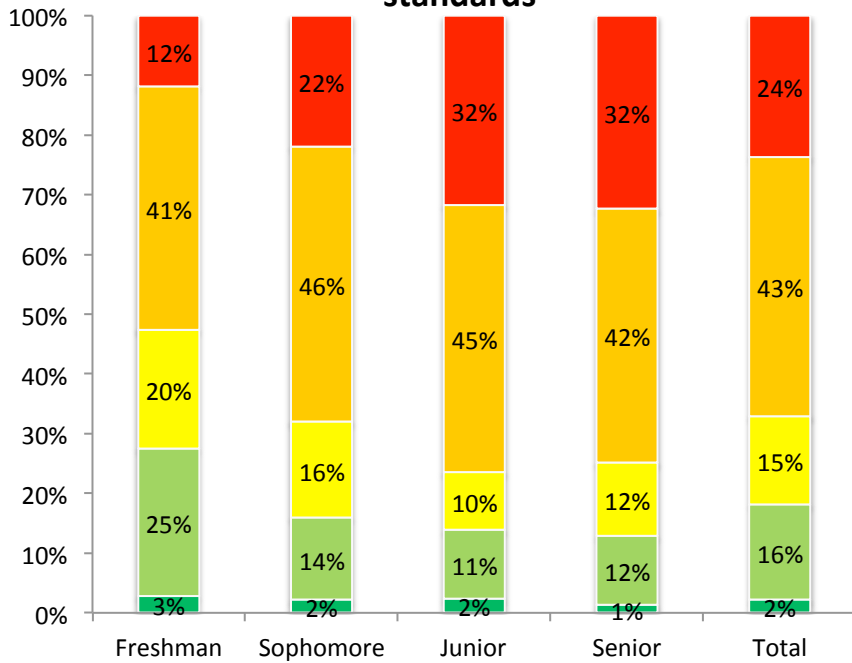


Q2B. How effective do you feel the grading policy has been in meeting the following objective?
 Grades give clear signals about the difference between ordinarily good work and the very best work.



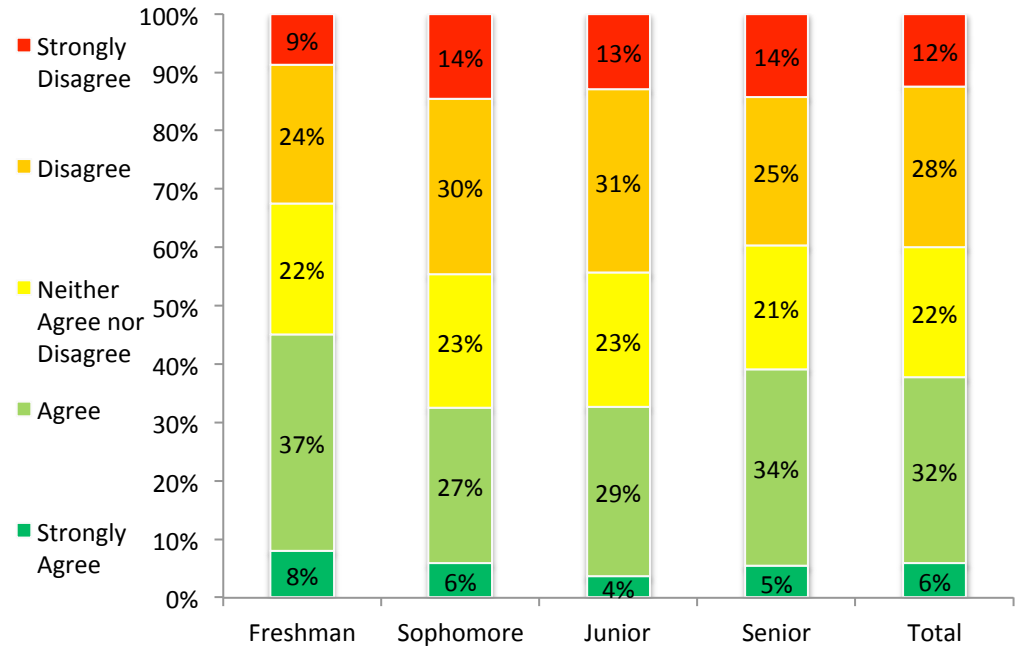
Q3A/B. STUDENTS: Based on your experience during your time at Princeton, to what extent do you agree with the following statements?

Across the courses I have taken, I have been graded according to the same standards



Count 692 616 543 553 2404

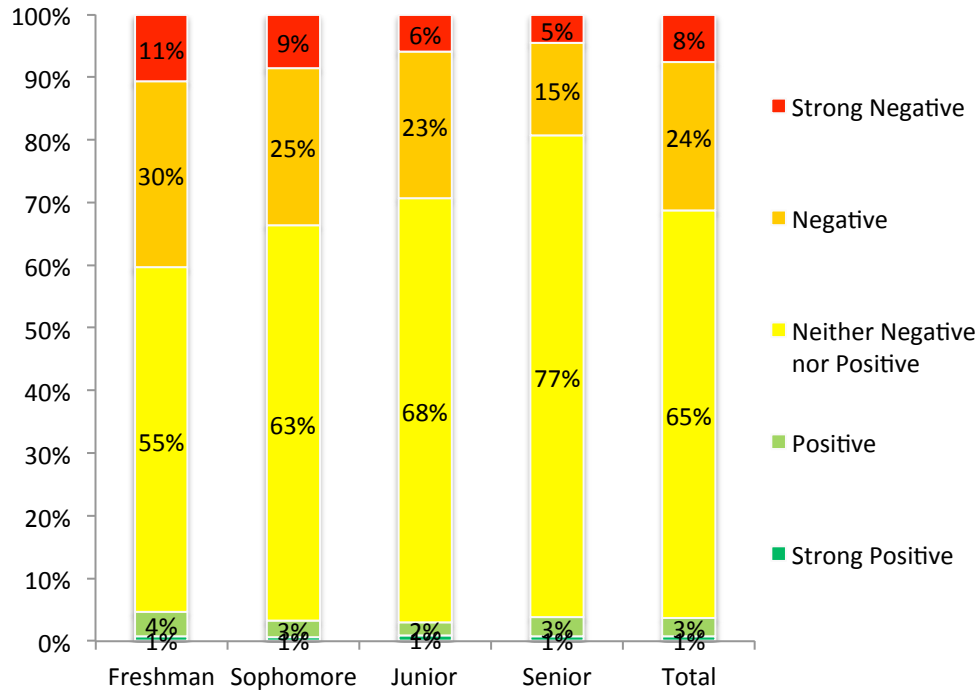
The grades in the courses I have taken have given me clear signals about the quality of my work



Count 691 614 541 554 2400

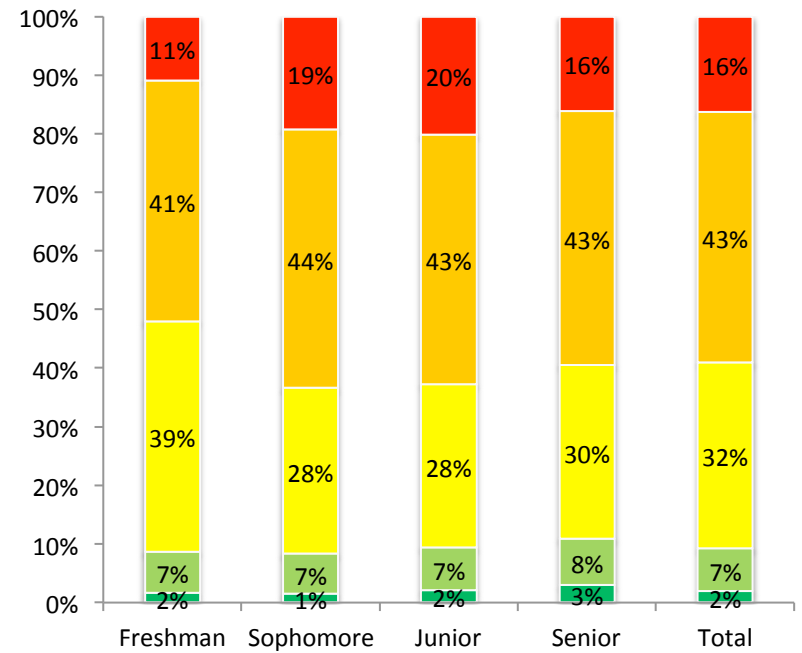
Q4/5. STUDENTS: Thinking about just your personal experience, how did the grading policy influence:

Your decision to come to Princeton



Count 696 619 545 550 2410

Your overall academic experience while at Princeton

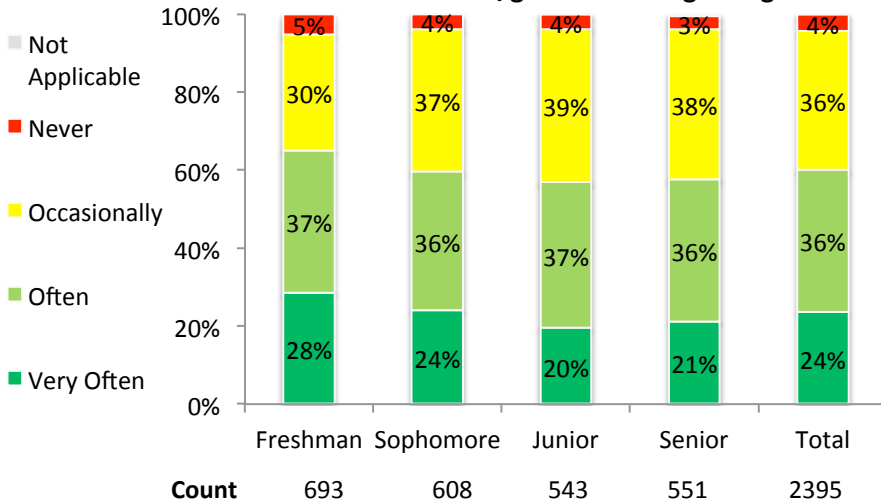


Count 696 618 545 552 2411

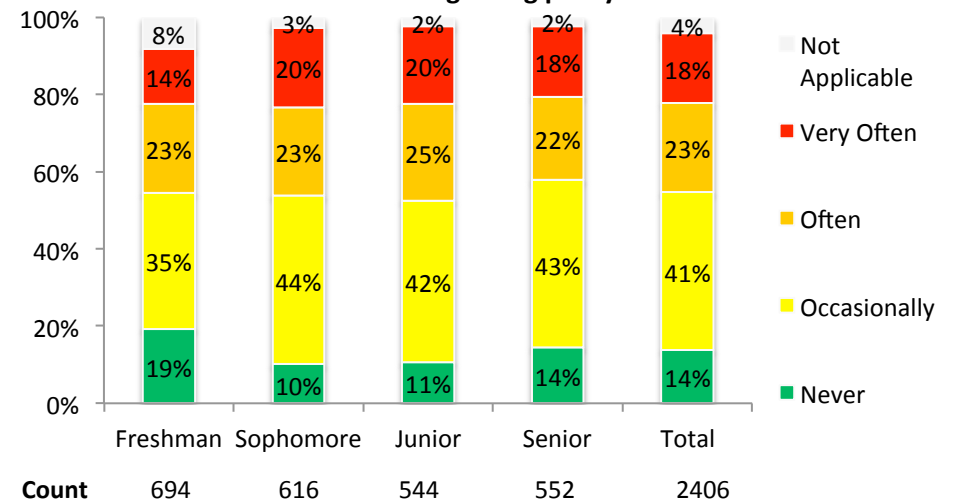
Q6. STUDENTS: During your time at Princeton, how often have your professors and/or other instructors (preceptors, lab instructors, language instructors, et al.) done each of the following?

** indicates that "Not Applicable" responses have not been reported.*

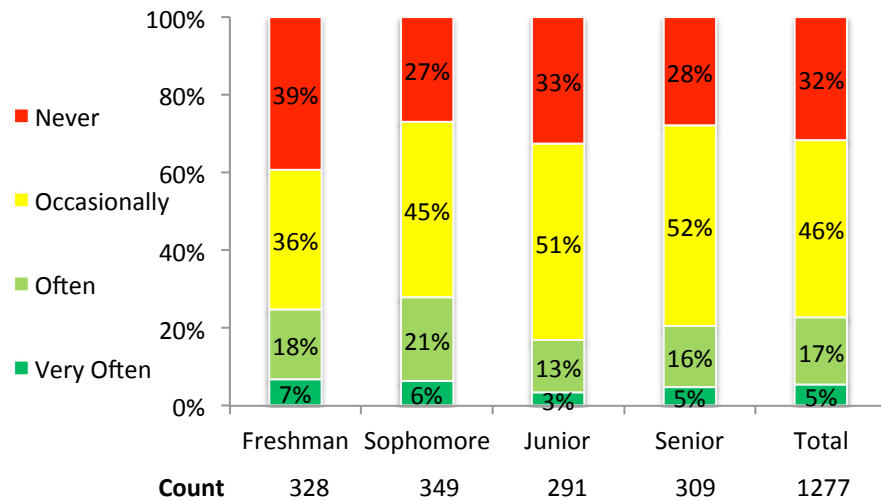
Discussed or provided their standards/guidelines for grading



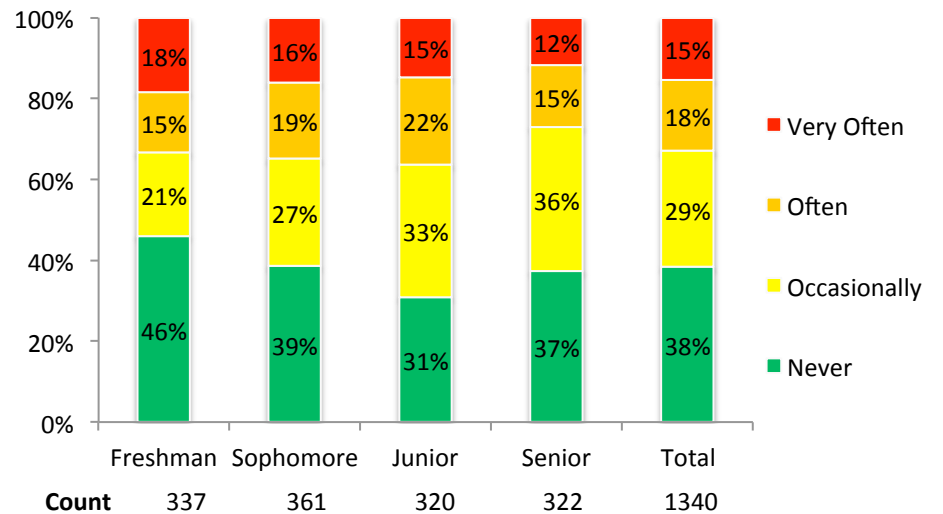
Given you a non-A course grade you felt was the result of the grading policy



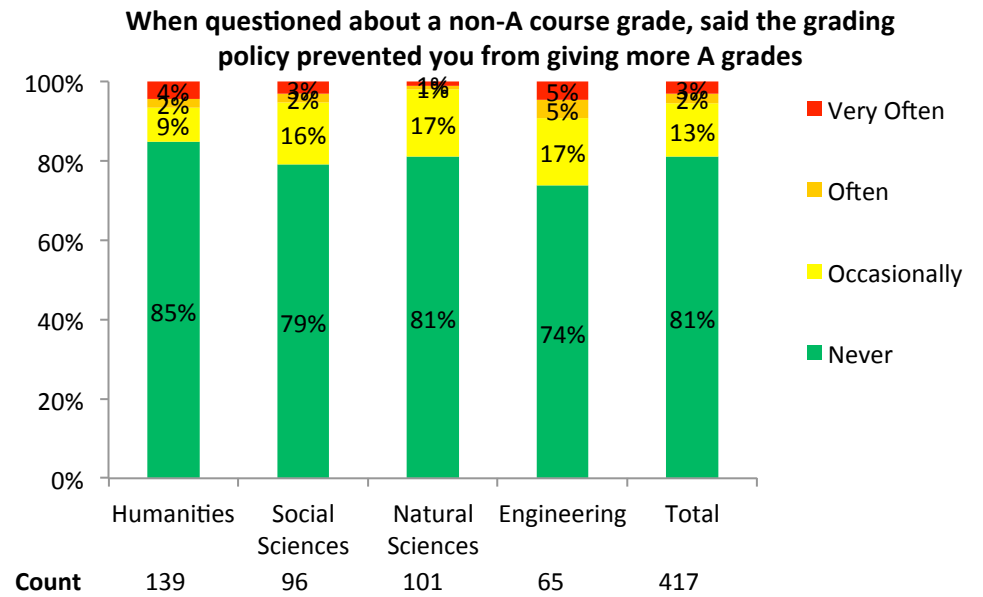
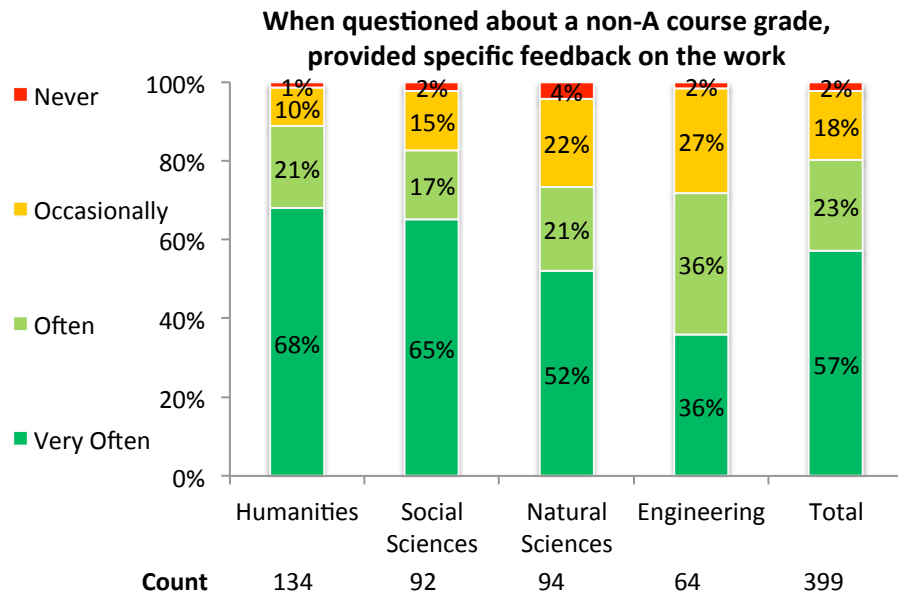
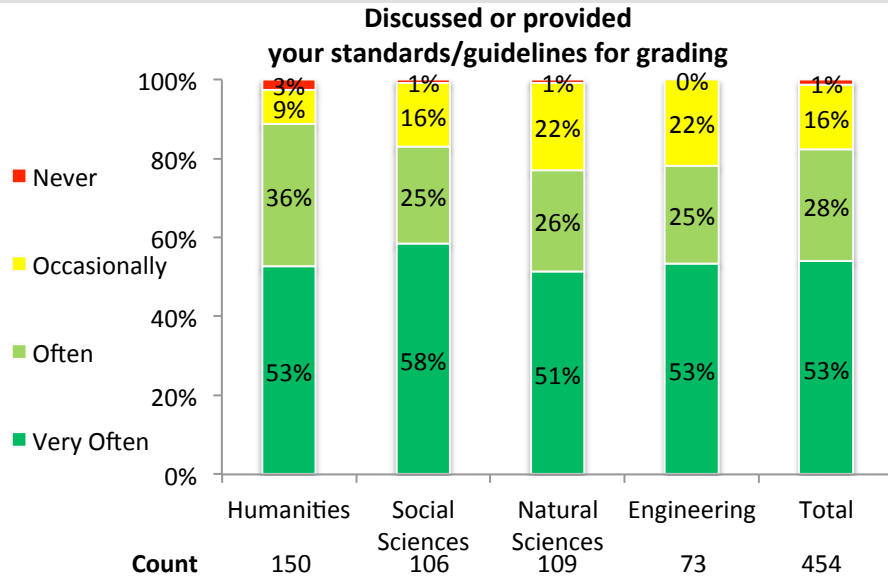
When questioned about a non-A course grade, provided you with specific feedback on your work in the course*



When questioned about a non-A course grade, said the grading policy prevented them from giving more A grades*

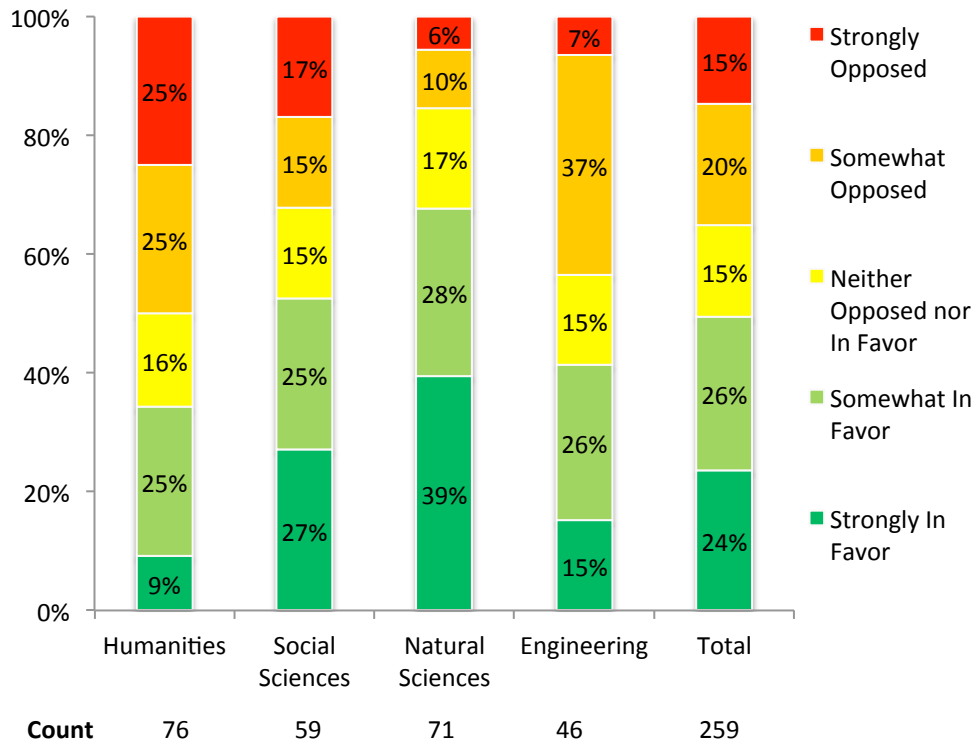


Q6. FACULTY: During your time at Princeton, how often have you done each of the following?
“Not Applicable” responses have not been reported.

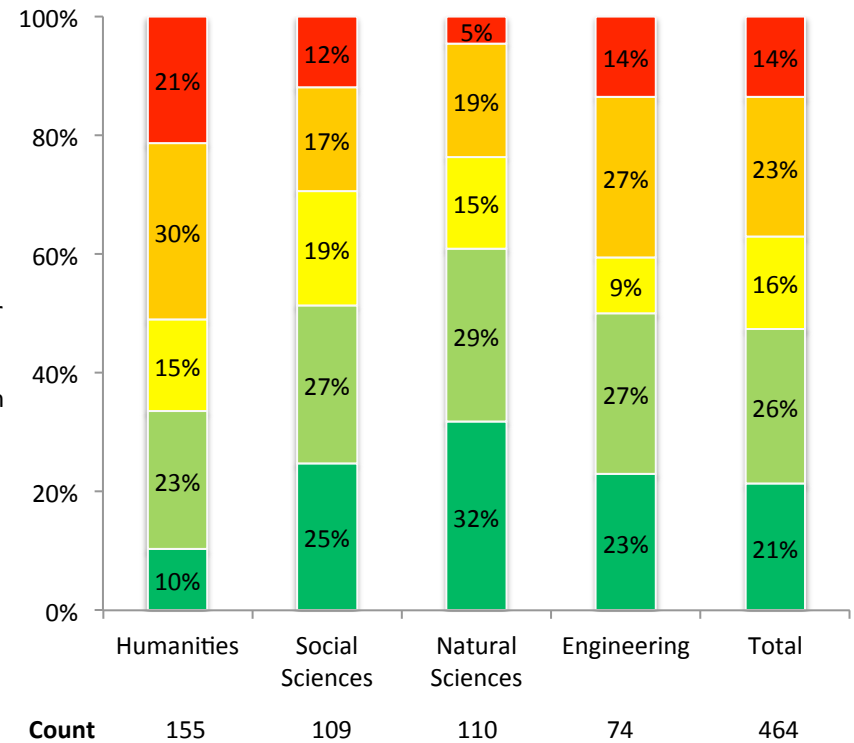


Q7. FACULTY: The current grading policy, which provides guidelines on the percentage of A grades awarded, was approved by the Faculty in April 2004.

What was your position on the grading policy in April 2004?



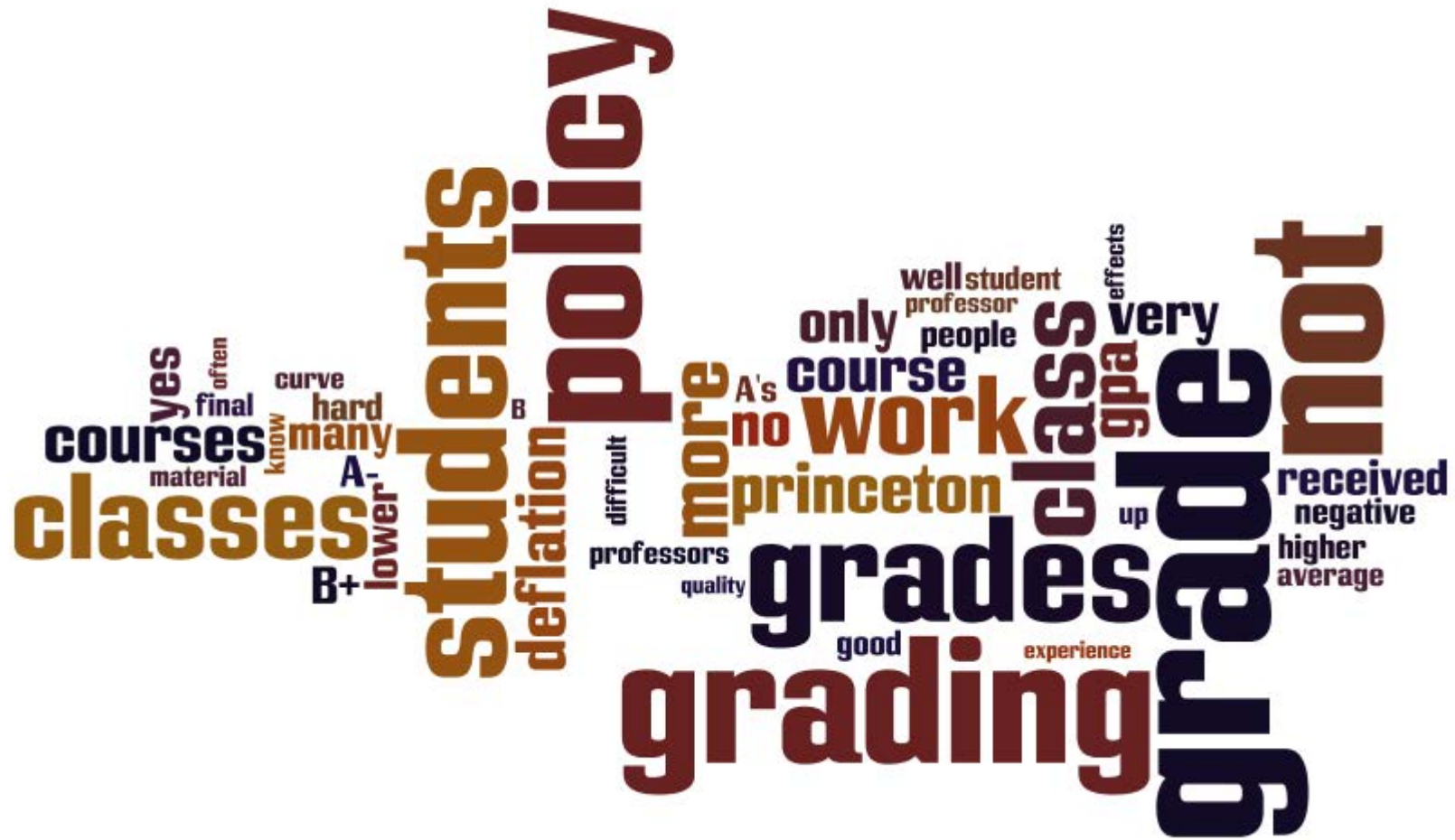
What is your position on the grading policy now?



FACULTY: Have you personally experienced any negative effects from the grading policy? Have you personally experienced any positive effects from the grading policy? Do you have specific suggestions for improving Princeton's grading policy?



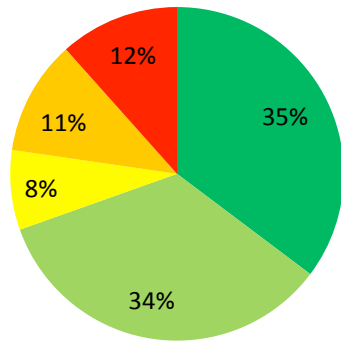
STUDENTS: Have you personally experienced any negative effects from the grading policy? Have you personally experienced any positive effects from the grading policy? Do you have specific suggestions for improving Princeton's grading policy?



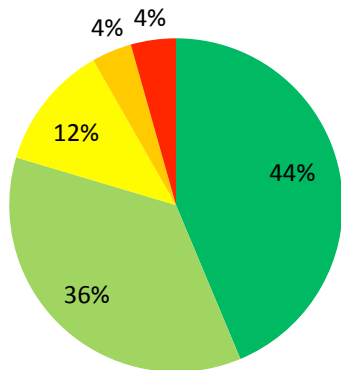
The public comment website received 211 submissions—half from undergraduate alumni, one-fifth from current undergraduates, and one-sixth from parents.

Goals

Students in one academic department should be graded according to the same standards as students in any other department.

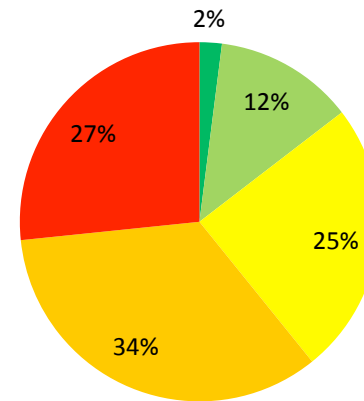


Grades should give clear signals about the difference between ordinarily good work and the very best work.



Effectiveness

Students in one academic department are graded according to the same standards as students in any other department.



Grades give clear signals about the difference between ordinarily good work and the very best work.

