

Revised Report of the Ad Hoc Committee on Grading

April 2013

1 Introduction

This committee was appointed by Dean Mary Miller in the fall of 2012 to look into Yale's undergraduate grading policies. We examined previous Yale reports on grading, collected data on Yale undergraduate grades over time and across departments, looked into grading practices at other institutions, and sampled opinions among our colleagues. Two of us reported preliminary findings to a Yale meeting of DUS's and met with students on the Dean's Advisory Committee to get their ideas, both in the fall of 2012.

A preliminary report of the committee was circulated and discussed at the February 7, 2013, Yale College Faculty meeting. Two of us met with students at a YCC meeting on March 27, 2013. Since the February 7th meeting, some departments have met to discuss the report and their grading policies, and the committee has received many comments and emails, both from faculty and students.

The committee first considered the general question of the purpose of grades. Why grade at all?

2 The Purpose of Grades

Grades serve educational purposes by signaling to a student strengths and weaknesses in performance. Grades on individual assignments give guidance as to what needs more attention, what can be used to build on, what needs change, and what bears repetition. They point to talents and capacities. Course grades signal possible directions for further study, for selecting majors, and for pursuing graduate or professional study. At the same time that grades give guidance, they also have a motivational function by encouraging, or not, activities of various sorts. They may elicit greater effort or discourage it.

Grades also serve the selecting and sorting functions of society by signaling to outside entities the strengths and weaknesses and overall capacity of individual students. They serve as certifications of competences, both particular and general.

Because grades are part of the teaching process—of the relationship between a teacher and his or her students—any suggestions about grading policies must be made carefully. The committee feels, as discussed below, that mandatory policies about grading are inappropriate, just as mandatory policies about many other aspects of teaching are inappropriate.

3 The Data

The committee next examined data on grades at Yale. Consider first Chart 1. It was created by the Office of Institutional Research (OIR). It is a remarkably informative chart in giving a history of grading at Yale. It shows that the percent of grades 90-100 in Yale College in 1963 was 10 percent. The percent of high grades (first 90-100, then Honors, then A) increased rapidly and roughly linearly until 1974, due in large part to the wider social revolution taking place in the United States at the time. In 1974 the percent was 40 percent. This percent stayed roughly constant until 1983, when it began to rise again. The last observation on the chart, for the spring of 2012, is about 62 percent.

Chart 1 also indicates other events related to grading. Grading units were changed in 1967, 1972, and 1981. In the fall of 1967, grading goes from numeric to Honors, High Pass, etc. In the fall of 1972 grading goes from Honors, High Pass, etc. to A, B, C, D, F. In the fall of 1981, pluses and minuses are added to the grading scale. Regarding other events, course evaluations began in the mid 1970's, the Course of Study Committee issued a grade report in May 1978, and the OIR reports on grade distributions across departments, which were issued to the faculty, ceased after the fall of 1982.

Table 1 shows grades by area, class size, and level for the 14 years between 1998 and 2012. This table (unlike Chart 1) excludes labs, independent study, and senior essay courses. The table shows that between 1998 and 2012 the mean course grade increased by 0.16. Remember that an A is 4.0 points, an A- is 3.67 points, a B+ is 3.33 points, and a B is 3.0 points. If we define a "unit" to be the distance between two consecutive grades, namely 0.33, then there has been a half a unit change in the last 14 years. The size of this change is roughly the same across departmental areas, the level of the class, and the size of the class. For all of Yale College the mean course grade was 3.57 in 2010–2011 and 3.58 in 2011–2012, rising from 3.42 in 1998–1999. Although not shown in the table, the percent A's and A-'s in the two academic years 2010–2012 was 61.8 percent. If classes with enrollments greater than 50 students are excluded, the percent rises to 66.8 percent,

with a mean course grade of 3.62.

Chart 2 presents similar data to those in Table 1 for the percent A's and A-'s, B+'s, B's, and B-'s, etc. These data are by semester.¹ Between the fall of 1998 and the spring of 2012, the percent A's and A-'s increased from 55 percent to 68 percent. One way of looking at the 68 percent number is the following. Consider the pool of students in the bottom half of the class. 36 percent of these students receive on average an A-.

The data in the two charts and in the table show that grade compression at Yale is quite high and that there is no indication that the trend to greater compression has ceased.

We also examined the data across departments for the two latest academic years: 2010–2011 and 2011–2012. We examined all departments that had enrolled 300 or more students, which totaled 51. These data are presented in the (unnumbered) table on the last page of this report. The table shows that the range across these 51 departments in the grading for these two years is large. The grades were for “standard classes,” which excludes labs, individual studies, and senior essays. Regarding the percent A's or A-'s, the range was from 47.7 percent to 82.5 percent. For mean course grades, the range was from 3.36 to 3.81.

It has been pointed out to the committee that the above range may be smaller than actually exists in practice. In less generous grading departments the fraction of students who drops a class for fear of getting a low grade is probably larger than it is in more generous grading departments, and the above range is only for students who complete the course.

Some suggested that these differences across departments may be due in part to some departments having more talented students. To examine this we collected for each department the average entering academic index² for the majors in that department. This index is to some extent a measure of academic quality, and it is not influenced by any decisions the students made while at Yale. (It is exogenous in this sense.) It turns out that there is a negative correlation between this index and a department's mean GPA (or percent A's and A-'s). The least generous grading departments have on average majors with higher academic indices. This analysis is very crude, but at least it does not suggest that grading differences across

¹The percents in Chart 2 exclude independent study, labs, and courses with less than 1 credit. Unlike in Table 1, they also pertain only to the population of students who got a grade of A through F. A grade of withdrawal, for example, is not counted for the percent calculations.

²The Academic Index, a measure created by the Ivy League to ensure that student athletes are “representative” of the undergraduate student bodies to which they are admitted, is an index comprised of the SAT I (or ACT), SAT II, and high school GPA.

departments are due to differences in average student quality.

The mean Yale GPA across all departments (not just the 51) was 3.57 in 2010–2011 and 3.58 in 2011–2012. How does this compare to other institutions? Not surprisingly, perhaps, it is very hard to get data on this. On the website called *gradeinflation.com* the estimated mean GPA at a selection of private schools, which includes Harvard and Princeton, but not Yale, is 3.30 for the 2006–2007 academic year. This compares to 3.51 for Yale in the same year using the data compiled for this report.

To summarize, grade compression appears to be continuing across almost all departments and the range of differences across departments continues to be large. For many departments now there are in effect only three grades used: A, A-, and B+. For the less generous departments, B is added to this group. Yale is approaching the point, at least in some departments, in which the only grades are A and A-, which is close to having no grading.

A final point about grades compressing at the top, the cutoff grade last year (2012) for *summa*, which is the top 5 percent of the students ranked by GPA, was 3.95, up 0.02 from 2010. For *magna*, the next 10 percent, the cutoff was 3.89 in 2012, up 0.04 from 2010. For *cum*, the next 15 percent, the cutoff was 3.80 in 2012, up 0.04 from 2010.

4 Costs of Grade Compression

The committee next considered the costs of grade compression at the top and the costs of large differences across departments.

- A. To the extent grades provide information, grade compression degrades that information. An excellent or outstanding performance is hard to flag if so many A's are given. High grades given for less worthy performances may mislead students as to where their strengths lie.
- B. Grade compression also muddies the effectiveness of grades as signals to outside organizations. Excellence may drown in the sea of abundant A's.
- C. Students may feel that their work is not taken seriously if most of the time they get at least an A-. Perhaps more seriously, students may put less effort into course work if it is easy to get a high grade. A recent NBER paper has data reporting that students worked an average of 40 hours in 1961 on class work and an average of 27 hours in 2003. Put another way, the reward for doing

excellent work versus good work may be less when grades do not distinguish between the two, and students may respond to this by simply working less on courses.

- D.** The cost of wide differences across departments is that they may distort student choices. If a student would really like to take class 1 over class 2 but chooses class 2 because it is graded more generously, this is a serious distortion. Even more serious is if a student would really like to major in A, but chooses B because the major grades more generously.
- E.** When grades are compressed, the inevitable noise in the grading systems, the random fluctuations, the effect of an individual instructor whose grading habits are significantly different from the average, either higher or lower, can have a much larger effect on relative grade point averages. Under Yale's current scheme, this can effect selection for graduation honors.

5 Proposals

It became clear to the committee fairly early on that many faculty have very little knowledge of grading practices in other departments and sometimes even in their own departments. It does not appear that many departments meet regularly to discuss grading policy. The first two of our proposals are meant to encourage more discussion and to provide more information to guide the discussions.

Proposal 1: Require the chair of each department to submit a written report to the dean of Yale College each year on the department's grading policies, with the expectation that departments will meet on a regular basis to discuss their grading practices and policies. Such discussions may be helpful in particular for new faculty. Graduate students who teach should also be brought into these discussions. Departments are also encouraged to include as part of this discussion how students are evaluated and given feedback.

Proposal 2: To facilitate the department discussions, return to making available to all FAS faculty the OIR grade report that ceased after the fall of 1982. This will make faculty aware of what other departments are doing, which may help them in formulating their own policies. The grade report gave distributions by department (not by individual classes). This report could be released at the end of each academic year by the registrar. The distributions by course in each department should be made available to the department chair. This should provide useful information in an individual department's discussions of its own grading policies.

If only these two proposals were adopted, this may just encourage a race to the bottom as less generous departments realize they are less generous and move to becoming more like the others. For example, a few years ago the economics department, which has recommended grading guidelines, realized that it was less generous than average and voted to make the guidelines more generous.

The committee thus strongly felt that more was needed than just proposals 1 and 2. The third proposal concerns changing the units of the currency. Studies of hyperinflations have shown that two things are necessary to stop them: structural reforms and changing the units of the currency. In the case of grades, hopes and habits seem to become fixed around certain letter grades, an "A" for example, and all they symbolize. Changing expectations will be much easier if these associations become less fixed. In order to defamiliarize everyone and start afresh,

we are proposing going back to numerical grading.

Proposal 3: Change to numerical grades, where passing is between 60 and 100 in units of 1 and 50 is fail. No grades are given between 51 and 59 and below 50. The Credit/D/Fail option would be changed to Credit/Fail, where fail is 50.

The shock value of moving from, say, 80 percent A's and A-'s to a much smaller percent is likely to be less if the grades are not in the same units. It is obvious there is a new regime. Numerical grades also have other advantages. There are no longer any "cliffs" between grades. If, say, only three grades are used, A, A-, and B+, the choice between an A and an A- or an A- and a B+ is of considerable consequence, which is not true if one can use many numbers. The consequences of the unavoidable randomness that occurs in any grading system are less serious with more grading choices.

Some have pointed out to the committee that students may argue more over one point than over an A- versus an A. This may be, but the consequences of a student winning or losing an argument are much less if it is just one or two points. If the consequences are smaller, maybe less effort rather than more will be devoted to arguing? It seems unlikely to the committee that this will be a major problem.

The committee has received many ideas about different grading units. Grading systems vary widely across U.S. universities and across the world. One of the cleverest ideas, from a Yale student, was to go to a Babylonian system of a 120 point scale. It divides neatly by both 3 and 4. In the end the committee felt that the most straightforward change was what has been proposed. It is easy to understand and stays within the realm of most people's experiences, while signaling a change of regime.

It has been suggested to the committee that some faculty may not be comfortable with so many grade choices, that some just think in terms of A, A-, B+, etc. It may be too fine a distinction to decide between, say, a 92 and a 93. If one feels this way, he or she can simply choose a few numbers to use and ignore the rest. For example, one might choose 97 for the old A, 93 for the old A-, 88 for the old B+, etc., and just use these numbers. This could be announced to students at the beginning of the term if desired. The guidelines discussed next could still be applied in this case, since they are in terms of numbers.

Some have suggested, particularly a number of students, that grading conveys more information if it is the form of written or oral comments from the teacher to the student. The committee agrees with this, but given the size of Yale and the need to convey information to the outside world, it is not a substitute for letter or

number grading. It is a compliment and should be done whenever feasible.

One way to alleviate student anxiety about grades in the new system and also to reduce haggling over minute portions of grades would be for instructors to distribute precise grading rubrics in each course. These rubrics would go beyond the usual percentage break-down of the overall course grade on the syllabus. Rubrics would indicate to students the exact elements of an assignment that add up to a particular grade. The committee believes that this would help specify, especially in non-quantitative fields, numerical equivalents to qualities of effort, insight, and execution.

On Credit/D/Fail versus Credit/Fail, last semester only one half of one percent of Yale grades were in the D category. So eliminating the D is trivial.

The committee from very early on was of the view that no mandatory grade distributions should be proposed. Each department and many courses have special circumstances, and requiring grading distributions would violate these special features. In addition, the committee feels it is presumptuous to dictate to any faculty member what he or she can or cannot do regarding grading.

The committee does feel, however, that suggested university-wide grade guidelines are useful for guiding a department's discussion of its grading policy. Suggested guidelines are in fact probably necessary for any meaningful change to take place. Without some frame of reference, it is unlikely that departments will all converge to similar grading practices. Similarity is particularly important if item D in Section 4 is to be addressed.

The following is the proposed set of guidelines. As Table 2 on the next page shows, these guidelines are consistent with the Mayhew memo of 1993, which sampled faculty opinion, and with the May 1978 report of the Course of Study Committee.

Proposal 4: Convey the following university-wide grade guidelines to the department chairs: 35 percent for 90–100, 40 percent for 80–89, 20 percent for 70–79, 4 or 5 percent for 60–69, and 0 or 1 percent for 50. If these guidelines were followed the mean grade would be 85.2. These percents are only suggestions and are in no way mandated. Each department has special circumstances. The guidelines are just that. They are meant to be used along with other information by the department in setting its grading policy.

Table 2
Grade Guidelines

Grade	Percent
Current Proposal	
100–90	35
89–80	40
79–70	20
69–60	4 or 5
50	0 or 1
Mayhew Memo (1993)	
A	34
B	43.5
C	18
D&F	4.5
May 1978 CSC	
A	40
B	37
C	11
D	2
F	2
Credit & Sat.	8

The May 1978 CSC report
thought the percents listed
were too generous and said
“some deflation is desirable.”

6 Further Discussion of the Proposals

The main point to realize is that the grading guidelines are not mandatory. Departments and individual faculty are free to do whatever they want. The hope is that faculty have a community spirit and try to do what is good for Yale overall, subject to their own individual circumstances. The registrar’s report if adopted would show to the rest of the faculty what the departments are doing, and departments that are way out of line may become more aware of how far out of line they are and try to

change over time, again subject to their own individual department circumstances.

To give an example of what we are not suggesting, say a department has a course that only its top students take, maybe one that has a number of difficult prerequisites. We are not suggesting that the 35/40/20/5 guidelines should be followed in this course. It may be that almost all students in the course get high grades, say in the 90's. Also, the distributions in large courses may differ from those in small courses. What we are suggesting is that at the end of the day, a department's overall distribution be close to the guidelines. But again, departments will differ somewhat, in part because they vary in the number of large classes they have.

Some information is available from the economics department as to the likelihood that the guidelines would be followed. The economics department has guidelines, and what happens each year is that the actual grade distributions are a little more generous than the guidelines, but each year there is a return to the base. In other words, the actual grade distributions are always a little more generous than the guidelines, but the amounts by which they are more generous do not change over time. If this were true for the university-wide guidelines, grade compression would be kept in check.

If the guidelines were roughly followed, the differences across departments would, of course, be lessened, which would lessen the distortions in students' decisions about what classes to take and what major to choose—item D in Section 4. There would also be, of course, no grade-compression problems, which would eliminate the costs listed in items A, B, C, and E in Section 4.

One might ask when we propose, say, 35 percent grades between 90 and 100, what reference group do we have in mind? If the reference group is all college students in the United States, 35 percent is very low given the selectivity of Yale. It is also likely too low if the reference group is Yale students from 80 years ago. The reference group we have in mind is current Yale students. Conditional on getting into Yale, how has the student performed relative to all others who got into Yale at the same time?

An interesting question is what has caused grade compression? It seems quite likely that the increase between 1963 and 1974 was due to the U.S. social movement. Student course evaluations, which began in the mid or late 1970's at Yale, probably have some effects on making grade compression worse. There is some literature on this. While this is interesting, it is of no help to the committee in designing proposals to change grading policies because almost no one would suggest eliminating course evaluations. (As an aside, a number of faculty have commented to us that there appears to be no grade compression in students grading faculty!) Eliminating the OIR grade report in 1982 probably had some effect on making

grade compression worse, which is why we are proposing going back to it.

7 Administration of the Changes

If the proposals are to be enacted, the new system would begin in the fall of 2014. The registrar would have general responsibility for implementing the changes and, along with Yale Career Services, reporting and explaining the changes to the outside world. Changing to a number system will make this reporting somewhat easier. Special reporting would be needed for students who were partly under the old system and partly under the new one. The committee has talked to the registrar about implementation, and the fall of 2014 looks feasible. The registrar would also be responsible for any changes on a student's transcript that would be helpful in communicating the changes.

The committee became aware as it did its homework that the system adjusts very well to different grading practices across universities. Take as an example applications to law schools. All applications first go to the Law School Admission Council (LSAC). The LSAC converts grades to their own 4 point scale and sends to the law schools 1) the student's GPA (in the 4 point scale), 2) the average GPA over the past three years of all the students from that student's university who applied to any law school, and 3) the percent ranking of the student relative to all the other students in the three year period (at the student's university). So the fact that Yale has a higher mean GPA than, say, Princeton, makes no difference. The differences are adjusted for in what seems to be a very accurate way. And when there is a transition period, as was true of Princeton some years ago, this is also known and adjusted for. Medical schools and business schools have similar organizations.

The committee did not consider the issue of departmental honors. If Yale's grading policies are changed, some committee, possibly a standing committee, will need to discuss this issue.

8 Relative Grading and Transcripts

The committee considered a number of options regarding relative grading and what should go transcripts. In the end these options were rejected. The following is a brief discussion of our reasoning.

Many ideas have been advanced, both at Yale and at other universities, for adjusting grades relative to something like the mean grade in the class. A student

getting an A in a class where the mean grade is an A maybe should count less than if the mean grade is a B+. The mean grade in a class could be recorded on the transcript and possibly even used to adjust the overall GPA.

These ideas are not, however, as fair as they may sound. Say there is an advanced class in a department that is generally taken by only the top majors in the department. Presumably most students in this class would get an A or an A- (or above 90 in the numbering system). It would be unfair to adjust the grade in this class down because all the grades were high. (This is also why we are not proposing guidelines for individual classes and are not making them mandatory.) It also has the potential for distorting choices of classes to take. Consider a top student who can get an A in any class he or she takes, and say that the student would like to take a class that has generous grading. The alternative is a less generous grading class, which the student is less interested in. In either case the student will get an A, but the student may be inclined to take the less desired, less generous grading class because the adjusted GPA will be better and/or the transcript will look better. It thus does not seem like a good idea to have a student's grade depend on the grades received by the other students in the class. (More advanced proposals have been made to adjust the adjustment by, say, the overall GPAs of the other students in the class, but this is likely to be an administrative nightmare and cause endless confusion.)

On transcripts, it turns out, much to the committee's initial surprise, that they are not used that much. Many potential schools and employers do not look at them. They are simply not a widespread way of communicating information to the outside world.

9 Student Reaction

As noted in the Introduction, we met with students on the Dean's Advisory Committee in the fall of 2012 and with the students at a YCC meeting on March 27, 2013. We have also received many emails from students and talked to many. A survey of students was taken in March 2013 by the YCC. Our impression is that many students, perhaps a majority, believe there is a problem regarding grade compression at the top, the continuing trend, and the wide differences across departments. There is, however, no agreed upon solution among the students as far as one can tell.

Regarding the survey, a student at the YCC meeting pointed out that it was framed in a very biased way, framed to get negative and almost panicked answers.

Also, there is considerable selection bias in the responses that were received. But for what they are worth, the results are as follows for four questions (the numbers are percents). 1) Do you believe Yale has grade inflation? Yes 44, no 35, uncertain 20. 2) Do you believe any changes need to be made to address the current grading system? Yes 25, no 59, uncertain 16. 3) Are you in favor of a 0-100 grading scale instead of letter grades? Yes 10, no 79, uncertain 11. 4) On the whole I believe effects of the proposed changes to grading grading will be: Positive 7, negative 79, uncertain 14.

A number of students have expressed the view that if there were a 100 point system, they would fret more about each point and generally have to work more on getting “good” grades than they now do. This would cut into their extra curricular activities, which are an important part their Yale experience and education. On the plus side regarding their emotions, however, is that if they have a bad day on a final and get, say, a course grade of 92 instead of 93, the consequences of this are much less than if under the current system they get a B+ rather than an A-. Bad draws have smaller consequences.

10 Previous Attempts at Yale to Change Grading Policy

An appendix to this report consists of three documents. One is a copy of the May 1978 Course of Study Committee (CSC) report on grading. The second is a copy of the 1980-1981 OIR report on grade distributions by department. The third is a memo dated October 22, 1993, by the then chair of the Course of Study Committee, David Mayhew, discussing recommended grading percents.

The May 1978 CSC report is interesting. Like the current report, it does not recommend mandatory guidelines. It felt that any further grade inflation (from 1978) would be “unconscionable,” and it hoped that the faculty would follow its advice and deflate. It relied on the OIR reports to provide information and guidance. Little did it know that the fall of 1982 would see the last OIR report and that between 1978 and 2012 the percent A’s and A-’s would rise from about 40 percent to over 60 percent.

It could be that the present report does no better than the 1978 report did (and now at a much higher grade level), but the committee feels that the chances for success are better. The proposed reporting process is more systematic; we are going back to the OIR report (hopefully to continue indefinitely); we are changing

the units of the currency; and we are recommending specific percents rather than just some general deflation.

11 Committee Members

Russell Adair, Office of Institutional Research

Suzanne Alonzo, Department of Ecology and Evolutionary Biology

Ronald Eyerman, Department of Sociology

Ray Fair, Department of Economics, Chair

Roger Howe, Department of Mathematics

Tina Lu, Department of East Asia Languages and Literatures

Paul North, Department of German

Mark Schenker, Dean of Academic Affairs

R. Shankar, Department of Physics