A proposal for updating grading policies

Executive Summary

The Subcommittee on Grading of the Undergraduate Educational Policy Committee recommends a system to address the external and internal challenges of grade compression:

- *A 20%+4 cap on A grades (but not on A– or other grades):* The Student Handbook recognizes an A grade as one reserved for work of "extraordinary distinction." We recommend returning to this definition and giving it a quantitative interpretation – a 20% cap – to harmonize top grades across courses and departments. Because small courses attract advanced and highly motivated students, we recommend allocating an additional four A grades to each class to raise the effective cap for smaller courses (for instance, a course of 10 students may allocate up to 2+4, or 60%, A's; a course of 100 may allocate 20+4, or 24%, A's).
- *Internal ranking by average percentile rank:* Letter grades compress information about relative student performance. We encourage instructors to submit raw scores along with letter grades, and recommend calculating internal honors using a student's average percentile rank (APR) rather than grade point average (GPA).

While any changes to grading policies may raise concerns about fostering a competitive culture, we believe that these recommendations take critical steps towards the College's goal to re-center academics, restoring confidence in the College's grading system, and better aligning incentives with pedagogical goals.

Background

In November 2024, Dean of Undergraduate Education Amanda Claybaugh appointed a committee to investigate grading policies and alternatives.[1] The problems with current grading practices are well-known, and rehearsed in Dean Claybaugh's memo Re-Centering Academics at Harvard College: Update on Grading and Workload.

The underlying problems with grading ramify in high levels of grade inflation. Of course, grade inflation at Harvard is not a new phenomenon; it has occurred more or less uniformly for over a

[1] Paulina Alberto, Professor of African and African American Studies and of History; Sarah Champlin-Scharff, Director of Academic Policy; Joshua Greene, Alfred and Rebecca Lin Professor of Civil Discourse; Adam Hesterberg, Associate Director of Undergraduate Studies in Computer Science; Alisha Holland, Gates Professor of Developing Societies; Raiyan Huq, Business Systems Analyst; Hanna Huh, Associate Registrar for Enrollment Services; Lisa Laskin, Assistant Dean for Undergraduate Education, Academic Program Support; Samantha Matherne (2024-25), Professor of Philosophy; Dominic Mao, Assistant Director of Undergraduate Studies, Molecular and Cellular Biology and Chemical and Physical Biology; Gillian Pierce, Associate Dean of Undergraduate Education, Academic Programs and Policy; Katherine Powers, Associate Director of Undergraduate Studies, Psychology; Stuart Shieber (chair), James O. Welch, Jr. and Virginia B. Welch Professor of Computer Science.

century. Over the last few years, however, what was a merely quantitative increase in average course grade has become a qualitative failure of the grading process as a whole. The increase in average grade has generated a compression of grades so pronounced that two-thirds of letter grades issued are straight A's and almost 85% are A-range grades (A or A–; see Figure 1).

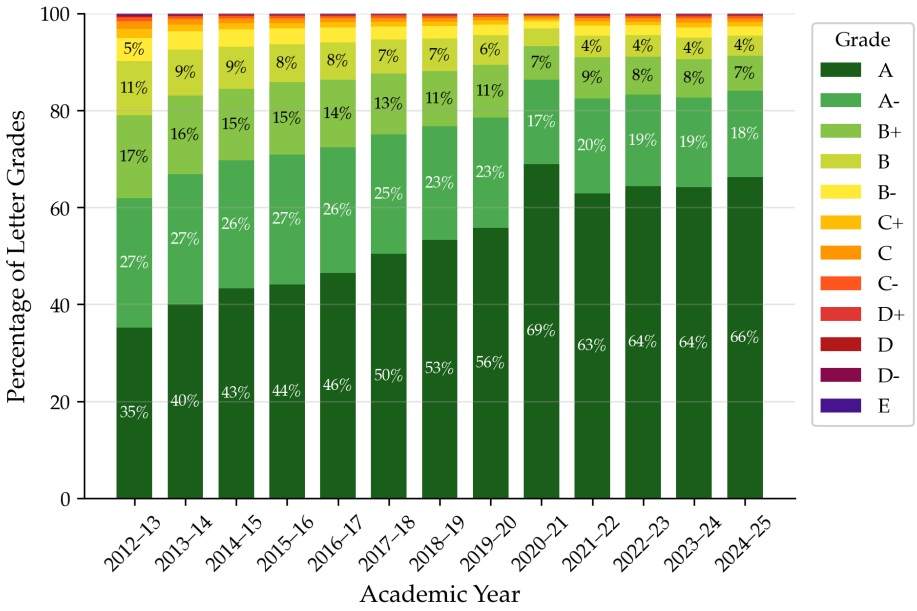

Figure 1: Grade distribution by year, from 2012-13 through 2024-25. (Fall and spring terms only. Partial data for 2012-2016.)

Grades as a measure of comparison are used for purposes both internal to the university – in determining honors, awards, fellowships, and the like – and externally – by potential employers, graduate admissions officers, and others with a legitimate interest in comparing student performance. As the Claybaugh memo states "We owe our students a functioning grading system. Specifically, we owe them grades that send clear signals, that give them a good sense of their strengths and weaknesses and that communicate their areas of distinction to employers and admissions committees."

Unfortunately, our current grading policies plainly do not support these roles for grades, both internally and externally. We propose a set of grading policies to mitigate these problems of grade inflation and compression for both the internal and external audiences.

The recommendations we present here are just one component of the broader effort to re-center academics in the College. In recent years, faculty committees and the Office of Undergraduate Education (OUE) have documented rising grade inflation and compression, uneven classroom engagement, and a tendency to prioritize extracurriculars over classes. In response, the OUE has encouraged individual faculty to review their own grading practices, departments and programs to articulate common standards, and the Faculty as a whole to consider policy changes that make grades more informative. The present recommendations extend this work by proposing a shared

framework for the use of high grades and the determination of internal honors. In doing so, we hope to clarify what the existing system of grades represents, calibrate grading across courses, and create a fairer process to distinguish students for honors.

Together, these initiatives are intended to restore grades to their role as meaningful indicators of student performance and feedback, and to support the central academic mission of Harvard College: teaching and learning. While this proposal focuses on limiting use of the highest grades, its spirit is to expand the potential of grades to communicate feedback and to enhance student learning. By encouraging faculty to use a wider spectrum of grades, we invite colleagues to design systems of assessment that align with their learning objectives and provide more frequent and better opportunities for detailed feedback on a student's mastery of skills or knowledge. The goal, in other words, is not to distinguish for the sake of creating distinctions, or to reward students who come to our classes already possessing the skills and knowledge that will allow them to distinguish themselves. Rather, we hope these initiatives will create space to use grades as pedagogical tools to lead students toward desired learning outcomes and to create the conditions for students willing to put in the effort to achieve mastery and even distinction over the course of their careers.

The current FAS grading rubric

Grades reflect student performance along multiple dimensions, most importantly, level of *mastery* of a body of knowledge and skills and relative *distinction* in applying the knowledge and skills. Mastery is an absolute measure whereas distinction is a relative measure.

Mastery is a precondition for distinction. Thus any grade conveying distinction should exceed whatever top grade(s) solely convey mastery.

> A, A– : Earned by work whose excellent quality indicates a full mastery of the subject and, in the case of the grade of A, is of extraordinary distinction.
>
> B+, B, B- : Earned by work that indicates a good comprehension of the course material, a good command of the skills needed to work with the course material, and the student's full engagement with the course requirements and activities.
>
> C+, C, C- : Earned by work that indicates an adequate and satisfactory comprehension of the course material and the skills needed to work with the course material and that indicates the student has met the basic requirements for completing assigned work and participating in class activities.
>
> D+, D, D- : Earned by work that is unsatisfactory but that indicates some minimal command of the course materials and some minimal participation in class activities that is worthy of course credit toward the degree.

E : Earned by work which is unsatisfactory and unworthy of course credit towards the degree.

Figure 2: The current FAS grading rubric.

The FAS already has a grading rubric (Figure 2) that reflects these principles directly. It uses grades A through E (with + and –) to convey levels of mastery, and the top grade of A unmodified to *further* convey distinction (in fact, "extraordinary distinction").

We support the FAS rubric in its structure of scaling letter grades with mastery and reserving the top grade for distinction. However, incentives in the system have led the rubric to be widely ignored. As a result, it serves poorly both the internal role for grades and the external role.

The internal role of grades

One role of grades is to provide information about students' relative performance for internal university purposes, including determining levels of honors, selecting recipients of awards and prizes, and establishing eligibility for fellowships and scholarships. Current practice relies on grade point averages (GPA) for comparing and distinguishing students' overall performance.

These uses only make sense if GPA is a meaningful signal of relative performance. But the quantity of grade inflation has in recent years qualitatively undermined the usefulness of grades for these purposes. As GPAs accumulate against the wall of 4.0, the small numerical differences that remain are less reflective of genuine variation in academic performance than random noise in the grading process. This compression is evident in phenomena such as the sharp post-2010 increase in Sophia Freund Prize winners (see Figure 3) and the shift of the *summa cum laude* cutoff so close to 4.0 that *summa* eligibility now depends on GPAs carried out to five decimal places. When honors and opportunities rest on distinctions this fragile, the results are not only unreliable but unjust, and continuing to use GPA as the primary metric for comparing students' academic performance is no longer defensible.

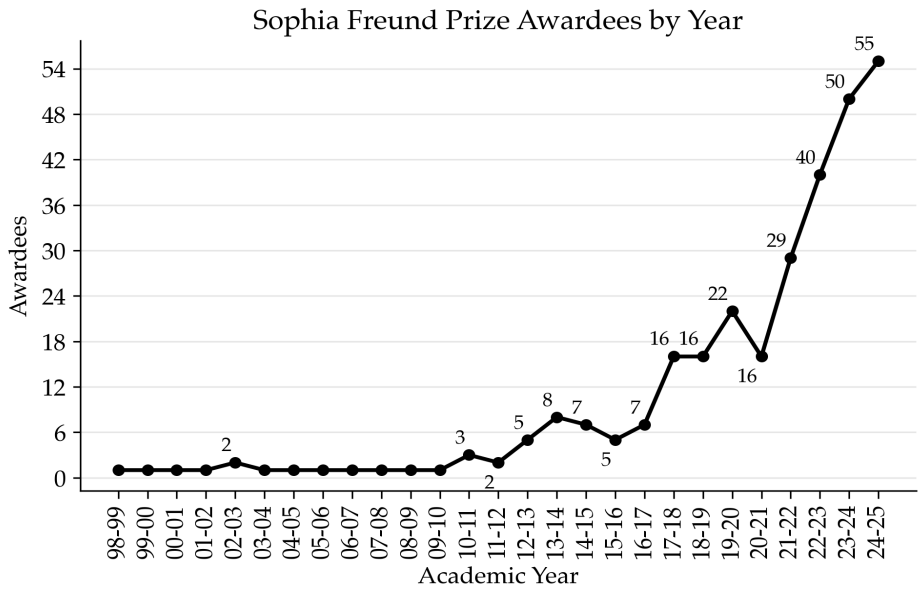

Figure 3: Superlinear growth in the number of Sophia Freund Prize awardees. The prize is awarded to "that student graduating summa cum laude who has the highest grade point average."

In order to base internal rankings on a metric that better captures genuine performance differences rather than chance variation, an alternative metric should be used that directly reflects those differences. A wide variety of such metrics are available and were considered by the committee. We've come to rest on a particular choice, and one that is particularly simple, *percentile rank*: one's percentile rank in a course is, informally speaking, the percentage of students in the course at or below you in ranking.[2] A percentile rank can be computed for each letter-graded course a student completes, and that rank averaged to calculate an average percentile rank (APR). The APR can take over the role of comparing student performance for internal purposes from the GPA metric that is manifestly dysfunctional for this purpose.

Percentile rank is especially attractive because it compares students directly to their peers rather than relying on fixed point scales or cutoffs. It remains stable even when assignments or courses differ in difficulty, and it is less sensitive to unusually high or low scores. It can be computed using only the letter grades of the students in the class, but can also take advantage of additional scoring or ranking information beyond that provided by relative letter grade alone. (See Recommendation 2.) In practice, this makes APR a clearer and more consistent indicator of how a student is performing relative to others, without requiring instructors to negotiate complicated scoring rules.

[2] More formally, a student's percentile rank in a course is $(r - 1) / (n - 1)$, where n is the number of students in the course and r is the student's rank from lowest grade ($r = 1$) to highest ($r = n$). When multiple students have identical grades, each receives the average of the ranks they would otherwise occupy. For example, two students tying for 5th and 6th place would each get rank $r = 5.5$ for the purpose of the percentile rank calculation.

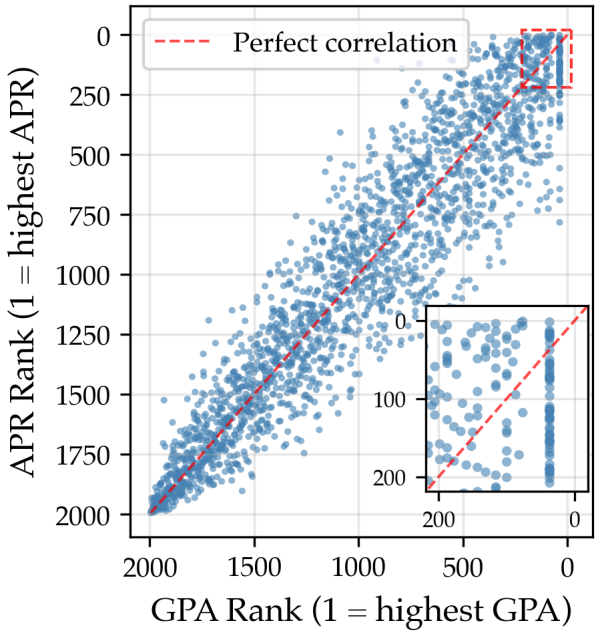

Figure 4: GPA rank versus APR rank for all students in the 2025 graduating class. The two metrics are highly correlated (about 94% rank correlation). The inset shows the students in the top 10% of both metrics. Note the striking artifact of a virtual column of students with identical GPA rank, caused by grade compression meeting the 4.0 barrier. This is the "Sophia Freund effect". No such artifact appears in the top APR ranking. That is, there is no similar row artifact.

Because percentile rank incorporates more information about students' relative performance in courses than the absolute letter grade, in empirical studies of their use we don't see the reflexes of grade compression so obvious from looking at GPA. There is no piling up at the top of the metric as there is with GPA. (See Figure 4.) There would have been, for instance, no growth, much less superlinear growth, in Sophia Freund prize winners when using APR.

> **Recommendation 1:** For internal purposes, the College should use average percentile rank (APR) rather than average grade (GPA) for quantitative comparison of student performance.

Although APR calculated over letter grades is a much better metric than GPA for evaluating relative performance of students for internal purposes, it can be improved further. As discussed so far it makes use only of the ordering of students by their letter grades in the course. But instructors typically have far more information about the relative performance of their students than that. Indeed, a typical approach instructors use in assigning letter grades is to first calculate some *score*, a weighted average, say, of the various assignments and evaluation opportunities for each student, and then to bucket these scores into the letter grades. The ordering defined by these scores is consistent with the ordering from the letter grades, but more fine-grained. Ideally, we should use the ranking of the underlying scores rather than the grades where possible.

Recommendation 2: Instructors should be allowed and encouraged to submit with their letter grades any scoring or ranking of the students consistent with the letter grades, to be used in calculating the percentile rank of the students in the course.

By replacing GPA with APR as the internal metric for relative performance, we would dramatically decrease the noise in our internal processes of rewarding high performance.

The external role of grades

We turn now to problems with grades in their role informing external audiences of student performance. A good example of the dysfunction of our current policies for the external audience, caused by our extraordinary level of grade compression, comes from the reports by admissions officers and employers that Harvard transcripts no longer provide them useful information about the performance and distinction of Harvard students, forcing them to rely on informal networks of information that advantage students with better networking connections. It should go without saying that privileged access to networks is not a fair basis for meting out jobs or grad school admissions.

The recommendations in the previous section were aimed at changing our internal practices so as to move from a quantitative metric awash in noise to one possessing more signal. However, it does nothing to address the underlying problem of compression of the grades available externally to the university, the grades as found on the transcript.

As noted above, the existing FAS rubric already provides for a wide range of grades for different levels of mastery and, in the case of the A grade, distinction. However, in evaluating "extraordinary distinction" the rubric provides no guidance as to what level or rarity of relative student performance qualifies as "extraordinary", and hence worthy of an A. Without such a recognized standard for "extraordinary distinction", instructors are left unmoored in their grading, and are subject to the multiple incentives to raise their proportion of A grades with no countervailing incentive to maintain A as a grade of distinction.

Consequently, over time the proportion of A grades has steadily increased, even without a concomitant increase in students' actual distinction in course work. Even if such an increase in performance had occurred over time, it would merely raise the bar for what ought to qualify as "extraordinary distinction" rather than increase its prevalence. Thus, explicit guidance to instructors on expectations regarding distinction (and hence A grades) would be helpful, and among many faculty, welcome.

Simply urging instructors to limit A grades to a particular proportion is unlikely to be effective in the long run. Doing so would put early adopters at a disadvantage when their courses compete with courses that continue giving higher grades. As long as grade inflation is widespread, no individual

instructor has a strong incentive to change, creating a collective-action problem. Many instructors report that they don't feel they can provide the differentiated grades that they think are appropriate for fear of dramatically reducing their enrollments.

Meaningful progress thus requires shared and coordinated policies, so that fairness is preserved for both students and instructors. We propose as a first step to limit the proportion of A grades awarded in a course to about 20%, consistent with the A rubric as characterizing "extraordinary distinction". Because this limitation on grade distributions differs from the policies of other universities, it is important that those interpreting grades be aware of the policy. It should be broadly advertised and accurately described on transcripts.

For very small courses of, say, a handful of students or fewer, a straight 20% cap on enrollment might allow for an unreasonably low number of students to receive A's, perhaps none. To allow for more flexibility, and recognizing that small courses often attract advanced and highly motivated students, we propose a lower bound of four on the number of awardable A grades by setting the cap at 20% of enrollment plus 4.[3] This cap on A grades has the property that large courses will be limited to about 20% A grades, while very small courses, small seminar courses for instance, will be able to award a much larger proportion of A grades. For instance, a course of five students would be able to award (5 × 0.2) + 4 = 5 A's, amounting to 100% of the class; a course of 10 students, 6 A's (60%); a course of 20 students, 8 A's (40%); a course of 100 students, 24 A's (24%); and so forth. Adding these extra few A grades per course would likely increase the overall percentage of A's from a nominal cap of 20% to an effective proportion of about 34%.

> **Recommendation 3:** Instructors may award A grades to at most 20% of the course enrollment plus 4. This policy should be broadly advertised, including on transcripts.

However, there may be instructors who believe that their pedagogy requires them not to limit themselves to the A cap. Allowing instructors to freely opt out of the A cap would, of course, reduce the cap to a mere recommendation, thereby reintroducing the collective action problem we are trying to solve, and returning to an inflated grade regimen with all of the signal-to-noise ratio problems we currently experience. For that reason, we recommend that a provision for opting out of the policy be carefully handled to maintain appropriate incentives. First, by opting out, an instructor violates the spirit of the "extraordinary distinction" rubric for A grades. It makes sense, then, that opted-out courses not participate in the calculation of the internal metric of performance (APR if Recommendation 1 is approved). In fairness to students selecting courses, if a course opts out, students ought to be apprised of this fact in a timely manner so that they can determine whether they want to enroll in a course whose grading practices will eliminate their grade from the internal calculations for honors, awards, and the like. Finally, in fairness to those interpreting transcripts, the fact that a course opted out of the cap should be available on the transcript. Fortunately, all of these

[3] We take the course enrollment to comprise undergraduate students enrolled for a letter grade at the end of term. This excludes students enrolled pass/fail or sat/unsat, or students who have dropped or withdrawn from the course. Currently about 60% of courses already comply with this cap, though that number has been decreasing over time.

aspects can be satisfied through the existing ability to administer a course under a SAT/UNSAT grading scheme.

> **Recommendation 4:** Courses may petition the OUE to opt out of the limit on A grades by a specified date, before the course registration period. The grading basis for courses that opt out will be SAT/UNSAT, and the courses will not participate in the calculation of the internal and external metrics of performance.

Finally, there may be cases where two courses provide more and less challenging versions of similar subject matter. In order for students to be no less likely to be awarded an A grade merely for their decision to take the more challenging course, it may be useful to allow the two courses to share a single cap. This can be implemented by transferring counts from the cap of one course to the other. In summary, our recommendation is

> **Recommendation 5:** By mutual agreement between two courses' instructors and with the permission of their department or area chair(s), in advance of the opt-out deadline, A grade counts can be transferred from one course's cap to the other.

A new grading culture

These recommendations are grouped together as a single proposal to balance some of the potential weaknesses of each reform. We see three synergies between the recommendations:

1) *Class Size Effects:* The 20%+4 proposal creates an incentive for students to take small classes with a higher effective cap on A's. Internal rankings, however, create the opposite incentive for risk-averse students. Small courses create a greater risk that a median grade produces a low average percentile rank. For instance, a student who receives an A– in a 10-person seminar may find themselves in the bottom 40% of the class. The higher effective cap on small classes also serves as a check on student "flocking" behavior towards "gem" classes. These incentives counterbalance each other, reducing the degree to which either change alone would incentivize students to game the system, and hopefully freeing students to select courses based on their interests more than the status quo does.
2) *The Top versus the Distribution:* The 20%+4 proposal applies only to A's, to recognize the Student Handbook's classification of an A as a distinction separate from all other mastery levels (that theoretically all students in a course can achieve). Nevertheless, internal ranking encourages instructors to consider the distinctions that they make among students and ideally the broader distribution of grades they use. While distinctions still can be created in a compressed range of raw grades, instructors who only enter letter grades better serve their students for honors by differentiating their relative performance. For instance, a student who receives an A– in a 100-person course where 20% of students receive A's and the rest receive A–s will be ranked about the 40th percentile. The same student who receives an A– in a course with 70% A grades will be ranked around the 15th percentile.

3) *The Top and the Very Top:* The 20%+4 recommendation interprets an A as a grade of extraordinary distinction, rather than adding an A+ to the College's grading scale. However, the internal ranking system allows instructors to make additional distinctions among A grades by entering raw scores (for instance, a 99 versus a 94), effectively allowing instructors to provide finer-grained information about the performance of their students.

Frequently Asked Questions

Details

Q: What about cross-registered courses? Will they count for the external or internal metric?

Harvard FAS already has a policy that courses whose grading scheme is indeterminate or independent of the FAS grading policies do not count in calculating aggregate metrics and are notated on the transcript as such. This policy applies to courses taken at MIT, study abroad courses, cross-registered courses offered by other Harvard faculties, and the like, and will apply to the new internal metric as well.

Q: I'm the instructor in a course that by virtue of its pedagogical structure cannot follow the 20% cap on A's. Can I opt out of this policy?

Harvard grading policy already has provision for courses that do not follow the standard letter-graded grading method. With OUE permission, you can specify that your course be graded SAT/UNSAT rather than for a letter grade, as is currently done for first-year seminars, junior and senior tutorials, reading and research courses, and certain introductory courses such as CS50 and Stat 110. The standard FAS policy will apply under which such courses are excluded from internal and external metrics and are noted as such on the transcript. In all such cases, information about the grading scheme will need to be made available to students by the start of course registration, so they have a chance to take the course's grading scheme into account.

Q: What about graduate courses?

The grading policies apply to all of the Harvard College undergraduates enrolled in a graduate course. The 20%+4 policy applies to the set of undergraduates, and the internal metric will be calculated based on the undergraduates in the course.

Q: What about graduate students?

The policy applies to Harvard College undergraduate students only. Over time, other schools, including GSAS, may institute similar policies for their students.

Q: As a faculty member, how do I calculate percentile rank? How will I submit this information?

Instructors don't need to calculate percentile rank. They will submit their letter grades to the registrar as before, optionally along with more detailed scoring information that they used to determine the letter grades. We envision doing so through an augmentation of the current grade submission system. The registrar's systems will calculate the percentile ranks. (For the formal definition of percentile rank, see Footnote 2.)

Q: When and how will the new system be implemented?

The cap on A grades will go into effect for the coming academic year. Upgrading the Registrar's systems to support all elements of the new grading system will require a phased approach over time and dedicated resources.

Q: How will these changes affect current students as we transition to the new grading policies?

Current students' grades will fall under the 20%+4 cap going forward. Thus, the constraint on A grades will affect rising seniors only for their senior year, rising juniors for their last two years, and so on, with incoming first-year students under the cap for all four years. For internal purposes, both GPA and APR metrics will be computed for all students so that either may be used by processes such as honors, prizes, and awards.

Faculty and student considerations

Q: Longer term, what will the effect of these grading policies be on students and on the reputation of a Harvard degree?

One important reason to address grade inflation at Harvard is to preserve the reputation of a Harvard undergraduate degree. Ultimately, it will benefit students if employers and graduate admissions committees know that the grades on a Harvard transcript are a more accurate and nuanced reflection of student performance in their courses.

Q: Will these policies place Harvard College graduates at a competitive disadvantage to graduates from peer institutions in graduate admissions or employment?

Harvard will widely publicize its grading scale and internal and external metrics for ranking and assessing students so that employers and admissions committees can accurately assess Harvard College graduates as measured against their peers graduating from institutions that have not, or not yet, taken similar measures to address grade inflation.

Currently, employers and admissions committees are routinely forced to consider information other than grades in their evaluation of candidates, since Harvard College grades are so undifferentiated. A recent report by FAS Institutional Research on grades and first destination outcomes for the past three graduating cohorts shows that employers consider a range of factors other than GPA, and

there is no substantial variation in outcomes to show that a high GPA is an absolute gatekeeper to elite employment, defined as highly competitive finance or consulting firms, highly competitive technology companies, other top employers of Harvard graduates, and graduate school admissions.

We know of no empirical evidence that students at other colleges that have tried related policies have been less successful in employment or admissions, including students at Princeton and Wellesley. Indeed, the grading cap may advantage Harvard students through the increased ability of employers and admissions committees to assess their abilities relative to others at peer institutions. An A at Harvard will provide more information than an A at, say, Yale, allowing external readers of transcripts more confidence in hiring or admitting the Harvard graduate.

A more articulated understanding of the likely effects on graduate admissions, employment, and fellowships is informed by discussions, summarized below, between representatives of the OUE and admissions deans and recruiters.

Medical and law schools admissions

We've spoken with the admissions deans at the law schools and medical schools to which our students most commonly apply. Their unanimous view is that grade inflation has become a serious problem. "The Harvard A doesn't make as much of an impression," a med school dean explained, "because there are so many." "It would be flippant to say that grades are useless," a law school dean observed, "but they're almost useless." Grade inflation is a problem in part because admissions committees are having trouble identifying the truly remarkable applicants in a sea of 4.0s. As a result, they must rely more heavily on the MCAT and LSAT than they'd like. But grade inflation is also a problem because a 4.0 isn't actually what admissions committees are looking for. They're looking for academically serious students who are willing to take risks and are able to rebound from setbacks, and they're willing to admit students with records that show that. (One law school dean noted that half the Harvard applicants admitted last year had at least one B+ on their transcripts). A 4.0, by contrast, can be the sign of a student who has always played it safe.

The admissions deans unanimously agreed that a 20%+4 cap would make our grades much more useful than they currently are. And they were confident that such a cap would not make our students less attractive applicants. "As we all know, grading practices are highly variable across institutions," one medical school dean pointed out, noting that requirements and the rules governing pass/fail and withdrawals are different at different schools, as well as the relative rigor of grading. "We in admissions are aware of all these factors and adjust accordingly." "We're still going to take Harvard students," a law school dean assured us. The admissions deans hope that we will make this change, and several predicted that other schools would follow our lead. "What Harvard does," one dean said, "sets the tone."

Employment prospects

As grades have risen, employers have given them less weight. National data show that only 42% of employers now screen for GPA (down from 73% prior to the pandemic). Internal data show that GPA is significantly less significant for Harvard students: of employers who post positions at Harvard, only 10.5% of internships and 3.0% of jobs set a minimum GPA. Many employers are relying on skills checks instead.

The employers who do set minimum GPAs tend to be boutique finance firms, not the larger firms that most students focus on. The six "top" finance and consulting firms, for instance, do not have minimum requirements, nor do the five "top" tech firms. Nonetheless, the six top finance and consulting firms do tend to hire students with GPAs that are somewhat higher than the GPAs of students who secure positions with other employers (3.89 median GPA vs. 3.82). The median GPA of students hired by the five top tech firms was 3.82. (It's worth remembering that, by definition, half the students who got jobs at these firms had GPAs *below* the median.) Last year, students with GPAs as low as 3.12 were hired by the six top finance and consulting firms. When asked about changing grading policies, recruiters asked only that we keep them apprised of the changes so they could adjust their expectations when reviewing applications.

Fellowships prospects

No internal Harvard fellowship specifies a grading minimum; some external fellowships do. The Rhodes and Marshall require a 3.7, and the Truman as well, but the Truman allows exceptions. The Goldwater and McCall MacBain require a 3.0. A 20% + 4 cap would not disqualify students from applying for any of these fellowships. Nor, our fellowships office believes, would it make our students less competitive compared to students from other schools, because fellowship committees assess academic performance holistically. Grades are important, but more important are candidate statements and interviews, as well as letters of recommendation, and Harvard students excel on these metrics.

Q: What are the implications for academic freedom and faculty autonomy?

In crafting this proposal, considerations of academic freedom and faculty autonomy weighed heavily. Our deliberations revealed that the critical question is not whether to prioritize the academic freedom of the faculty, but how best to balance competing academic freedoms.

Under the current system, a great many faculty feel unfree. They feel unable to give grades that accurately reflect the quality of student's work, denying students useful feedback and opportunities to distinguish themselves in the classroom. As students exercise their freedoms to choose their courses and evaluate their instructors, many faculty – and especially junior faculty and non-ladder faculty – feel pressure to give higher grades, yielding dramatic grade inflation. Thus, under the current system, grading decisions are free in theory, but not in practice.

The present proposal aims to solve this problem using the least restrictive means that is nevertheless up to the task. The proposed policy does not tell instructors how to teach, what assignments to give, which students should receive which grades, or which ideas may be expressed in class. To the consternation of some faculty members, the present proposal grants individual faculty the freedom to give all A-range grades. Faculty who teach courses with pedagogical goals that conflict with the cap, such as community-based learning experiences or tutorials that emphasize collaboration, can opt out of the system, as described in Recommendation 4. And this proposal recognizes that a high proportion of students in small, upper-level courses can produce work worthy of distinction. This proposal simply aims to preserve the A as a meaningful mark of distinction, bringing grades in line with the spirit of the FAS grading rubric.

There are clear precedents for the Faculty, collectively, to place some limits on individual course decisions. We already accept constraints on when courses can meet, whether courses must be taught in person, and the range of grades allowed (no A+, for example). Likewise, we have collectively agreed on a standardized rubric for what letter grades should mean. The proposed cap is another such shared framework designed to promote more effective teaching and learning.

The reason for a common framework is that grades function as a shared signal. A faculty-adopted cap on A grades is a solution to an urgent collective action problem. This proposal aims to counteract perverse individual incentives to inflate grades while preserving instructors' core freedoms to educate students as they see fit and to the best of their abilities.

Q: Will students from under-resourced high schools be disproportionately harmed by the new policies?

A key function of grades is to provide feedback on learning. All students, regardless of background, should receive accurate signals about their strengths and areas for improvement to guide course and concentration choice. FAS Institutional Research data show that students from under-resourced high schools do not currently perform lower in GPA to any significant degree, so we do not expect the new policies to affect them disproportionately.

All the admissions deans and recruiters we talked to were quick to recognize the value of admitting students and hiring employees who bring varied life experiences. They recognize that some students are more likely to struggle in the transition to college, and they are much more interested in the trend in grades over time than in the total GPA.

Q: Wouldn't these policies introduce competition among students for the "scarce resource" of A's, leading to increased stress on an already overstressed population?

Competition is inevitable in any evaluative system where distinctions matter. When grades are compressed to the degree we currently see, that competition gets displaced elsewhere, in activities outside the classroom. And the current situation of grade inflation produces its own stress, as a single A– grade is seen to harm a student's chances of graduating with honors or of being elected to Phi Beta Kappa.

By capping A's, we increase confidence that grade distinctions reflect both degrees of mastery and distinction. At the same time, instructors can provide finer-grained raw scores for use in internal evaluations, to prevent tiny differences from causing outsized anxiety. Together, these changes ground the notion of distinction, increasing the signal provided by evaluations. Instructors can go even further to reduce stress: by designing collaboration-friendly assignments, competition inside the classroom becomes healthier. And by reducing grade compression, we not only clarify distinctions but also align students' focus on meaningful, shared academic goals.

Q: Won't these recommendations just shift "shopping" behavior from looking for grading gems to looking for workload gems?

Perhaps, but even that is progress. Faculty may choose to design their courses to ensure that students are adequately challenged and assign work outside of class to meet accreditation standards, and the FAS will continue to monitor data on course workloads.

Q: This policy doesn't go far enough. What about capping the proportion of A– grades or B+ grades or the course GPA overall?

Only the A grade in the current rubric embodies *distinction* as opposed to *mastery*. A robust notion of mastery (as found in the descriptions of the other grades) is consistent with the possibility that all students may reach it, whereas any reasonable notion of "distinction" would involve some notion of distinguishing a select group.

For instructors looking for a model or target grade distribution beyond the A grade, one can be extrapolated as follows: If we think of the A grade as 20% of the class, and we target an A–/B+ mean grade, a clipped normal distribution yields the percentages shown in Figure 5:

Grade	**Percentage**
A	20%
A–	30%
B+	30%
B and below	20%

Figure 5: A possible target grade distribution for advisory purposes.

This distribution turns out to be roughly comparable to those from the early 2010s.

Q: Given the benefits of the internal percentile rank metric, why not place it on the transcript instead of or in addition to the letter grade?

The APR metric is intended to improve information about our own students for our internal purposes, and isn't appropriate for comparing across institutions that don't provide similar metrics. Over time, if the approach is picked up by our peer institutions, it may be appropriate to provide this information on transcripts as well.

Q: Will the change from GPA to APR affect some students more than others?

The APR metric, though highly correlated with GPA, is not identical. Some students' APR will be higher than their GPA, and some lower. (See Figure 4.) A worry is that the change is systematic. For example, perhaps students in humanities concentrations might be overrepresented in the top 5% of GPAs but underrepresented in the top 5% of APRs, or vice versa. If the former, the change from GPA to APR for the purpose of *summa* eligibility would reduce the relative proportion of humanities students eligible for that honor. The change to APR might be considered unfair. Or perhaps the status quo ante proportion under the GPA metric had been the unfair approach, which would now be corrected. Adjudicating this difference – which of the two was the unfair distribution – would require some notion of ground truth to compare against, which we don't have.

In any case, in empirical testing, though there are differences in proportion of this sort among the divisions, they change substantially from year to year in ways that seem random rather than systematic. We were able to find no consistent trends in students from different divisional cohorts being over- or underrepresented with the two metrics.

Alternatives considered

Q: Why not allow instructors to opt out of the 20%+4 cap while maintaining letter grades?

An alternative opt-out policy would be to allow faculty to give letter grades, uncapped, but without counting toward summary metrics (GPA, APR) and with a notation on the transcript. While this would grant an additional degree of freedom to faculty offering opt-out courses, this would create perverse incentives for both students and instructors, likely leading to inflated grades and confusing, uninformative transcripts.

Under such a policy, students could devote extra effort to a limited number of capped courses – in the limiting case, a single course – to secure a high GPA and APR. They could seek uncapped courses to build a "moat" around their towering GPAs and ranks, which are misleadingly based on distinguished performance in a small number of capped courses. The demand for low-risk, uncapped courses from students seeking to protect their summary metrics would pressure faculty to offer more uncapped courses, possibly making uncapped courses the norm rather than the exception. The result would be meaningless internal rankings based on a small subset of courses and a proliferation of transcripts reporting abundant A grades and extremely high GPAs. We recognize that, under the recommended policy, students could likewise attempt to protect their

summary metrics by taking a higher proportion of SAT/UNSAT courses. However, the prospect of a transcript filled with SAT grades is likely less appealing to students than one filled with As, even if many As are marked as products of uncapped grading.

We therefore recommend the familiar SAT/UNSAT as the grading basis for opt-out courses. Faculty using this grading basis are, of course, still free to provide their students with more fine-grained feedback and provide more detailed evaluations in recommendation letters.

Q: What metrics other than percentile rank were considered for internal purposes, and why were they not chosen?

There are other metrics that might be considered beyond percentile rank, for instance, z-score normalization (number of standard deviations above the course mean) or delta grading (difference between the grade and the course mean or median). A good sign of the appropriateness of percentile rank is that it is (perhaps unsurprisingly) *extraordinarily* highly correlated with z-score and delta grades[4] while being natural and relatively easy to understand. For those interested in experimenting with alternative metrics, a demonstration of some of the alternatives is available at http://shbr.link/metrics.

Q: Why not introduce an A+ rather than capping As?

The Student Handbook already establishes an A as a grade of "extraordinary distinction." We want to restore meaning in our already shared grading standards. Shifting the entire grading scale up to accommodate a new top distinction could create continued cycles of upward pressure and reduce the legitimacy of our current grading rubric, much as Spinal Tap tries to increase the volume by adding an "11" on the volume button. Moreover, proposals to create a capped A+ grade were even more restrictive (often discussing caps of less than 5 percent), which would only foster competition over an even smaller number of top grades.

Q: What is the likely alternative to passing this grading policy?

If the policy is not adopted, the current situation is unlikely to remain stable. Two developments are especially important.

First, it is highly likely that within the next five to ten years, elite institutions, including Harvard, will be under increasing pressure to revise grading practices. Grade compression has made it harder for parents, employers, and graduate and professional schools to interpret transcripts as accurate signals of student performance. In that environment, peers will move to restore clearer distinctions in grading. The main question for Harvard is whether to move now, on our own terms, or to follow later after other institutions have already acted and framed the standards of "rigor."

[4] Spearman correlations among them are in the 0.98-0.99 range, much higher than the correlations between GPA and any of these alternatives, around 0.94.

Second, our existing honors and awards system is on track to become untenable. As an example, the faculty have already voted to limit the proportion of students who may receive *summa* to 5%. At the same time, the proportion of students with 4.0 GPAs continues to rise. On the current trajectory, we will soon face a situation in which there are more 4.0 students than available *summa* slots.

At that point, the Faculty will be forced into one of a small number of unattractive choices: inflate *summa* to encompass nearly all 4.0 students; impose stringent grading limits later, under time pressure and with fewer options; or substantially scale back or dismantle the existing honors structure. Similar problems accrue to other honors and awards.

In short, the realistic alternative to adopting a clear, prospective grading policy now is not the preservation of the status quo, but a future in which Harvard must make sharper and more disruptive adjustments under less favorable conditions.

Previously tried policies

Q: How does this policy differ from grade inflation policies instituted at other colleges like Princeton and Wellesley?

To our knowledge, no college has used any of these particular methods: capping high grades, incorporating additional scoring information, or using an alternative metric to GPA.

In 2004, Princeton introduced a target on A-range grades (A+ through A–), monitored on multi-year departmental averages. The targets thus fall somewhere between a course cap (as in the present policy) and an exhortation. This had the benefit of allowing flexibility in reallocating high grades among courses on similar topics, a benefit we pursue through Recommendation 5. On the other hand, by distributing the cap over many classes and multiple years, responsibility for maintaining the cap becomes diffuse and ensuring compliance becomes more difficult. The Princeton policy was lifted in 2014.

By contrast, Wellesley in 2004 specified a grading constraint at the individual course level, but placed the constraint on the course's overall GPA rather than A grades in particular: courses with over 10 students were to have a grade mean not to exceed B+. This leads to the unfortunate incentive for instructors to "buy" more A's by lowering other student's grades.

The A grade cap has the benefit that the constraint manifests the natural notion of awarding distinction in student performance.

Some programs simply provide contextual information on the transcript, such as course median (Dartmouth, Cornell University) or median and enrollment information (UNC Chapel Hill). Adding course medians to the transcript would not be particularly informative to those interpreting the transcripts. As of academic year 2025, some 73% of enrollments were in courses with an A median, and 95% with an A-range median. Noting this on the transcript would merely emphasize the inflated range of grades, while providing a convenient source of information for students to identify

the higher median courses. More precise contextual information (such as course percentile rank based on detailed scoring) might be informative. Adding this information might be considered as we gain experience with the new grading policies.

Feb 5, 2026