HARVARD
COLLEGE

Re-Centering Academics at Harvard College: Update on Grading and Workload

October 2025

Amanda Claybaugh, Dean of Undergraduate Education

on behalf of the Office of Undergraduate Education

Table of Contents

Introduction

It's become a familiar ritual. Every May, the FAS Faculty assemble to vote on the various degrees and the Dean of the College announces the cutoff for summa that year. The number invariably prompts murmuring in the audience. In 2022, it was 3.976. In 2023, 3.986. The number often inspires faculty to speak. Some rise to deplore the inflation in our grading, while others, more historically minded, observe that it has ever been thus. The meeting then adjourns, and the ritual repeats the following year. In 2024, the *summa* cutoff was 3.985; in 2025, 3.989.[1]

But another number from this past spring points in another direction. In 2025, the number of students finishing their first year with a 4.0 was 251, down from 286 the previous year.[2] That's a small decrease, to be sure, but it reminds us that we're not, in fact, at the mercy of inexorable trends, that the grades we give don't always have to rise.

This report makes the case for further change. It begins by asking whether our grading is actually a problem. Our mission, after all, is to educate our students, not to grade them, and so our grading practices are a problem only to the extent that they interfere with that mission. The first section of this report argues that they do. It shows that our current practices are not only failing to perform the key functions of grading; they are also damaging the academic culture of the College more generally. The report then asks how we got to this point. Critics of higher education like to blame grading trends on snowflake students and negligent faculty, but that hardly describes the faculty and students at Harvard. Our grading trends are driven by other forces, and the second section of this report catalogues those. Finally, the report asks how we might reverse these trends. Other schools have tried and failed, and Harvard has as well. The history of these efforts reveals that exhortations won't be enough, nor is there a single policy fix. But coordinated action—individually, collectively, and institutionally—can restore the integrity of our grading and return the academic culture of the College to what it was in the recent past.

I. The Current Situation

It's true that grades always seem to be rising. It's also true that the rise has become extreme in recent years. Individual courses may vary, but most follow the same pattern: a slow rise in the early 2010s, continuous with longstanding trends, followed by a more rapid rise in the

[1] Minutes of the FAS Faculty Meetings, Office of the Secretary.
[2] Report from the FAS Prize Office.

late 2010s, then an additional spike during the year of remote instruction and a flattening out after that.[3] The result has been a difference of degree significant enough to become a difference in kind. We now have considerable compression at the top of the scale: in 2005, As accounted for 24% of all grades awarded in Harvard College; in 2015, for 40.3%; in 2025, for 60.2%.[4] And unless we believe that the work done by that 60.2% is truly of "exceptional quality" (the definition of an A in the *FAS Grading Guidelines*), we can see inflation in our grades as well.[5]

To get a more nuanced sense of what it's like to grade and be graded at Harvard College, the Office of Undergraduate Education spent the summer surveying faculty and students. We focused on people who were qualified to speak not only for themselves but also on behalf of others: the directors of undergraduate studies; the faculty teaching large introductory courses; the resident deans; the peer advising fellows; the students serving on the academic sub-committee of the Harvard Undergraduate Association. These surveys were complemented by discussions with the Educational Policy Committee, the Council of Chairs, and the OUE Student Advisory Council, as well as by informal conversations with many students and faculty. Everyone is eager to talk about grading.

When asked about grading in general, nearly all faculty expressed serious concern. They perceive there to be a misalignment between the grades awarded and the quality of student work. Faculty newly arrived at Harvard are surprised at how leniently our courses are graded, and those who have taught here for a long time are struck by the difference from the recent past. Students, for their part, were more sanguine. "I think the current grading system is working very well," one reported, "I haven't heard many complaints from other students either." Some students do complain, however, about a different kind of misalignment. They perceive there to be disparities in grading between concentrations, between courses, and even between sections of the same course. A preliminary look at the data suggests that they may be right.

When asked about the specific functions of grading, faculty and student responses reveal that those functions are not being fulfilled:

> *Motivation*: We expect grading to encourage students to engage with their courses. But faculty perceive—and students confirm—that in our current system grades are as much a limit on effort as a spur. Students will do what they need to do to earn an A, but they are

[3] Grading trends for individual courses can be found on the *FAS Grading Distribution Dashboard* under the tab "Compare Courses."
[4] FAS Office of Institutional Research, *Percentage of A Grades, 2006-2025*. Reprinted in Appendix.
[5] "FAS Grading Guidelines," *Harvard College Student Handbook* 2024-25, p. 29. Reprinted in Appendix.

seldom moved to do more than that. Once the A is secured, students direct their energies elsewhere, either adding more courses to their schedules or taking on more demanding extracurriculars.[6] What does encourage students to engage is their sense that a course is meaningful. "I knew what I was getting myself into when I chose this concentration," one student explained, "but I am so passionate about it that I am committed to it." Our current grading practices seem at best orthogonal (at worst, antithetical) to such meaning.

Information: We expect grades to give students reliable information about whether they have mastered skills or material, and students mostly believe that they do. But faculty report that the grades they're currently giving aren't differentiated enough for that. In the view of faculty, grades currently distinguish between work that meets expectations or fails to meet expectations, but beyond that grades don't distinguish much at all. "Students know that an 'A' can be awarded," one faculty member observed, "for anything from outstanding work to reasonably satisfactory work. It's a farce." Or, as another faculty member put it more delicately: "There's not much resolving power at the top." As a result, students don't always have an accurate sense of how they are performing or where their strengths truly lie.

Distinction: We rely on grades to distinguish the strongest student work for the purposes of honors, prizes, and applications to professional and graduate schools. But the grades we award no longer do so. People who have served on prize committees, for instance, find it hard to distinguish between students with identical GPAs. For some prizes, they rely on confidential letters to learn which of the many students awarded As actually distinguished themselves. For others, they must trust their own sense of which transcripts display a more or less demanding array of courses. At best, this system is opaque to students; at worst, it risks introducing bias and inconsistency into the process. What is true of prizes is true of graduate and professional school admissions as well. Increasingly, admissions committees tell us that it can be difficult to tell Harvard students apart.

Our current grading practices are not only undermining the functions of grading; they are also damaging the academic culture of the College more generally, by:

Constraining student choice: When the median grade is an A, a 4.0 is no longer a goal that a small number of students are pursuing, but rather a default that many students are trying to maintain. "We are terrified of the A-," one student confided, and so they act in the ways that research on loss aversion would predict. While their choice of concentration

[6] Students are taking more courses than they used to: where 82.64% of students enrolled in 4 courses a semester in 2015-16, only 72.63% did so in 2024-25; nearly all the rest enrolled in 5, 6, or even 7. FAS Institutional Research, *Courses Per Semester, 15-16 to 24-25*, slide 2. Reprinted in Appendix.

is driven by their sense of what would be most interesting and most useful, their choice of courses is instead determined by their sense of the grades they will receive. They feel compelled to choose courses that seem to guarantee As over courses they know would be more valuable for them or simply more interesting. And the more students make this prudential calculation, the more their GPAs converge. Aberrant grades become ever more costly.

Exacerbating stress: When the median grade is an A, earning an A is no longer enough. Students sense that their grades are insufficient to distinguish them from one another, and so they seek to distinguish themselves in other ways instead. Some are driven to pursue additional academic credentials, such as double concentrations, secondary fields, language citations, or a concurrent master's degree. Others focus on pre-professional clubs to the exclusion of activities that they might find more meaningful or enjoyable. In this way, students' stress about achievement expands beyond the classroom and pervades all domains of college life.

Hollowing out academics: When the median grade is an A, academics can start to feel, as students sometimes put it, "fake." Our students know what real achievement feels like: they experience it in their research, their sports, their clubs, in the start-ups they launch, and the non-profits they found. But they don't always feel that same sense of achievement in their courses. One student lamented that no instructor had ever told her that she could do better work, and several mentioned sitting for finals that they could have aced on the first day of class. The "gems" they take prove to be counterfeit.

Workload is notoriously difficult to measure, but our data suggest that students are working as hard as they ever have—if not more. In the spring of 2025, students reported on the *Q* that they spent 6.30 hours outside of class for each of their courses. In the spring of 2015, they reported spending 5.55.[7] College surveys point in the same direction. In 2025, the OUE survey asked students how much time they spent every week on academics: 45 hours a week was the modal response, compared to 10-15 hours a week on extracurriculars. In 2024, the Dean of Students Office asked students how much time they spent on extracurriculars: 83% of students reported spending less than 10 hours a week.[8]

Nonetheless, many faculty believe that hours worked have gone down, even as grades have gone up. Indeed, 69% of the faculty who responded to the *Harvard Crimson* survey this year either strongly agreed or somewhat agreed that "Harvard students do not sufficiently

[7] FAS Office of Institutional Research, *Q Score Trends Fall 2005 – Spring 2025*, p. 4. Reprinted in Appendix.
[8] Report from the Dean of Students Office.

prioritize their coursework."[9] How representative these faculty might be remains an open question, since the *Crimson* reveals little about its respondents, but the OUE's surveying of faculty reveals that the ones who are most concerned about workload are the ones who teach in reading-intensive disciplines. Faculty in the humanities and interpretive social sciences report that they've had to trim some readings and drop others entirely, that they've had to switch from novels to short stories, and that it's difficult to keep assigning reading in the face of increasing student complaints. Faculty in p-set disciplines report no analogous concerns.

A fair number of students in reading-intensive courses report doing lower than the average hours of work outside of class, and they might well be encouraged to do more by comprehensive final exams and reading quizzes. But another factor also seems to be in play. Students are coming to us—because of changes in high school education, because of changes in our media culture—with less experience reading complex prose and less capacity for focus and sustained attention. (Notably, results from the NEAP just revealed that reading scores for high school seniors in 2024 were down from 2015).[10] Little surprise then that students struggle with readings that students completed with ease just ten years ago, or that faculty perceive them to be putting in less time than they are. This is a much harder problem to address, but it's crucial that we find a way to prepare our students to grapple with the kinds of challenging texts that are central to so many disciplines—and to life beyond Harvard.

We don't have a problem with workload overall. Our accreditation requires that each 4-credit course entails 12 hours of work, and, in general, our courses are not falling far short of that. (The ones that are falling far short should either increase the work they are assigning or do more to encourage students to complete the assigned work). We do, however, have a problem with grading. Our grading is too compressed and too inflated, as nearly all faculty recognize; it is also too inconsistent, as students have observed. More importantly, our grading no longer performs its primary functions and is undermining our academic mission.

II. How Did We Get Here?

Faculty lament that our grading practices are, as one puts it, "out of whack." But many feel powerless to grade otherwise. About half of the faculty surveyed reported that they simply cannot award the grades students have actually earned, while the rest reported that they can

[9] William C. Mao and Veronica H. Paulus, "On Survey Majority of FAS Faculty Say Harvard Undergrads Don't Care Enough About Their Courses," *Harvard Crimson.* 2 September 2025. https://www.thecrimson.com/article/2025/9/2/faculty-concerned-student-priorities/

[10] National Center for Education Statistics, *2024 National Assessment of Educational Progress(NEAP) Reading Assessment at Grade 12,* 2024, *https*://www.nationsreportcard.gov/reports/reading/2024/g12/.

do so only with difficulty. Doing so requires "burdensome material preparation and rhetorical positioning," one faculty member reported, while another said that it requires considerable "strength of character." A third noted that it felt possible only after tenure.

What makes grading so difficult? Some of the pressure comes from other faculty. Faculty don't grade in isolation: they know that their assignments will be perceived as demanding and their grading as tough not on their own terms, but in relation to what other faculty are doing. No one wants to be an outlier. Faculty also know that they are competing with one another for students. With good enrollments come guaranteed teaching for graduate students, the possibility of more faculty lines for departments, and the opportunity to introduce students to the subjects we've devoted our lives to studying. It makes sense, in this context, for faculty to lower standards in order to raise enrollments. "Grading at Harvard is in a race to the bottom," one faculty member observed. "This is a classic game theory problem."

This classic problem seems to have intensified in recent years, following a change in course evaluations. Traditionally, these evaluations were fairly informal. The student members of the Committee on Undergraduate Education would distribute evaluation forms, compile the results, and publish them as the *CUE Guide*; faculty could opt out of evaluation if they chose. Then, in 2008, the FAS Faculty voted to formalize the process: Institutional Research took charge, faculty were required to participate, and the *CUE Guide* became the *Q*. Newly formalized, the scores given by students seemed to carry more weight. Instructors now worry a great deal about the *Q*. Teaching Fellows and non-ladder faculty fear that low scores will limit their job prospects, while tenure-track faculty fear that low scores will damage their tenure chances. Nor are tenured faculty immune, since they feel responsible for drawing undergraduates into their concentrations. These concerns shape grading. Even though grading has only a small predictive power for *Q* scores and workload has no predictive power at all, faculty nonetheless believe that the grades they award and the work they assign determine the *Q* scores they receive.[11]

Students also exert pressure on grading.[12] Sometimes, the pressure is direct, as when some students challenge their grades or take an "increasingly litigious" approach to every assignment. At other times, the pressure is indirect, with some advisors advocating inappropriately on students' behalf. Driving this pressure are very real—if exaggerated—concerns. On the one hand, students are right that their grades will play a role in determining whether they are admitted to medical school or law school or get a desirable internship; on

[11] Nina Zipser and Lisa Mincielli, "Do Difficulty, Workload, and Grades Correlate with Instructor *Q* Scores?," 3 March 2025. Reprinted in Appendix.

[12] It's worth noting that students come to us accustomed to earning very high grades. The Class of 2029 graduated from high school with a median GPA of 3.89.

the other hand, they tend to overestimate what GPA is actually required for these goals and underestimate the number of paths open to them. Their strong, if only partially accurate, fears make it difficult for faculty to award the grades that have been earned.

Another kind of pressure comes from the College itself. For the past decade or so, the College has been exhorting faculty to remember that some students arrive less prepared for college than others, that some are struggling with difficult family situations or other challenges, that many are struggling with imposter syndrome—and nearly all are suffering from stress. Faculty have taken these exhortations to heart. Unsure how best to support their students, many have simply become more lenient. Requirements were relaxed, and grades were raised, particularly in the year of remote instruction. This leniency, while well-intentioned, has had pernicious effects. Not only has it contributed to grade compression and inflation; it has also undermined our educational mission. There's been what one faculty member describes as "a radical shift from providing critical feedback to providing emotional support." Many faculty would like to reverse that shift; they want to return to doing the hard, but necessary, work of providing real feedback. They fear, though, that students would complain, and they wonder whether the administration would, as several put it, "have their back."

A final factor influencing grading is a new approach to teaching. Almost twenty years ago, the *Compact to Enhance Teaching and Learning at Harvard* (2007) called on the FAS to "reward excellence in teaching and advising" and to "improve pedagogy and student learning."[13] In the years since then, faculty have more than met this call. We have focused more attention on our teaching, and many of us have redesigned our courses to increase learning. Some of these changes, however, have had unintended consequences. For instance, many of us shifted from high-stakes exams to more frequent lower-stakes assignments, believing that this would help students retain the material. A number have found, however, that lower-stakes assignments are more effective at rewarding effort than at evaluating performance, giving students the false sense that they'd mastered material that still eludes them. Similarly, faculty shifted from exams and papers to alternate modes of assessment, such as creative assignments and group projects, in the hopes of increasing student engagement with their courses. A number struggled, however, to assess these assignments in a sufficiently differentiated way. Finally, some faculty have eschewed conventional grading, turning instead to "ungrading" or "contract-based learning" or other systems in which students earn As for completing all assigned work. There is a pedagogical case to be made for these alternate approaches, but they're fundamentally at odds with our current grading system, which relies on grades to differentiate.

[13] Andrew Biewener et al., *A Compact to Enhance Teaching and Learning at Harvard*, 2007.

Students shouldn't be blamed for pursuing As so avidly, nor faculty for awarding them so frequently. Both groups are operating reasonably within a system structured by competing incentives—and a culture animated by evolving beliefs about what students need and what good teaching entails. We must change that system and culture if we want grading to change.

III. What Should We Do?

We've been seeking to re-center academics ever since the pandemic. The first three years of this work focused on determining the extent and nature of the challenges we face. In the spring of 2023, an OUE committee issued the *Report on Grading in Harvard College*, which showed that our grades had gotten higher and more compressed.[14] In the spring of 2025, an FAS committee issued the *Classroom Social Compact Committee Report*, which showed that some students were censoring themselves in our classrooms and otherwise disengaging from their coursework.[15] Finally, this report has gathered more data about grading and workload, as well as surveying students and faculty to better understand the forces influencing both. No doubt, there is still more to understand, and the OUE is eager to explore further questions as they emerge.

But the time for concerted action has come. Some faculty have already taken steps to bring their own grading back into integrity, and others should now follow. We recommend that all faculty do the following before the spring semester begins:

> *Review the distribution of grades in their courses over time*. Faculty should reflect on how their courses were graded—or how courses like their courses were graded—in the recent past. (The *FAS Grading Distribution Dashboard* presents this information, under the tab "compare courses.") Our grading was not overly stringent in 2015, but we did assign a broader distribution of grades. That prior distribution might be a useful target for us now.
>
> *Review their assignments and the relative weight of each*. Faculty should consider instituting seated exams if they haven't already done so. Seated exams are prudent in this age of Generative AI. They are also useful for encouraging students to engage with all course materials, and they tend to produce a broader distribution of grades. Faculty should also reflect on how much weight they put on effort (attendance, completion of

[14] Gillian Pierce et al., Office of Undergraduate Education, *Report on Grading at Harvard College*. 17 February 2023. https://ouedocumentlibrary.fas.harvard.edu/grading-report.
[15] Maya Jasanoff, David Laibson, et al., *Classroom Social Compact Committee Report*. 2025.

problem sets) as opposed to mastery and determine whether effort-based grading is causing grades to rise higher than is warranted.

Articulate their grading standards. Faculty should explain to students what each grade represents. Our students don't know what constitutes "excellent" or "extraordinary" work in a discipline. Indeed, they may not even know that the quality of the work is what grades are evaluating. Certainly, the students surveyed for this report almost invariably spoke of grading in terms of effort. Grades feel fair to them, they said, when they work hard and get an A, while grades feel unfair when they work hard and don't get an A or when another student doesn't work hard and does. Only one student spoke of grades in terms of demonstrating skills or mastering material. There's clearly a need for faculty to be more explicit about what grades are actually measuring.

Calibrate grading across sections of a course. Students are troubled by inconsistency in grading, and the least we can do is make sure that grading is consistent within a single course. Course heads are responsible for ensuring that it is. Some do this by having TFs grade exams collectively, each taking responsibility for certain questions; others, by having TFs norm their grading on papers before assigning final grades. In addition to calibrating grading, strategies like these also insulate TFs from the complaints of students disappointed by their grades.

There are also steps that faculty can take collectively. Some departments have already come together to articulate grading standards for their concentrations, while the faculty teaching large gateway courses have met to discuss their grading as well. Other groups might follow their example: perhaps all faculty teaching creative courses, or all faculty teaching language courses, or all faculty overseeing junior tutorials. The OUE would be very happy to convene these conversations, if that would be useful.

Faculty can also act collectively by making new policies. The Educational Policy Committee has charged a faculty committee with exploring possible adjustments to our current system of honors and grading. This committee is open to all suggestions; at the moment, they're focusing on the following possibilities:

Allowing faculty to award a limited number of A+s. Currently, we permit faculty to award B+s and C+s, but not A+s. Permitting faculty to award a limited number of A+s in each course would increase the information our grades provide by distinguishing the very best students.

Recording the median grade for every course on the transcript. Currently, we provide no context for the grades a student receives, relying on rumors about which courses are graded more or less strictly than average. Recording the median grade for each course would provide useful contextual information. It would also reduce the pressure students currently feel to take easily graded courses, since the benefits of doing so would be less.

Creating a variance-based grading system for internal use. Currently, we use the transcripts prepared for external audiences for internal purposes as well (PBK, honors, other prizes). The committee is currently exploring alternate systems for expressing grades internally and modeling the effects of those systems.

The EPC committee will be seeking feedback on these proposals from students and faculty alike, and none will be adopted without the vote of the FAS Faculty.

Finally, there are steps that FAS and College leadership must take to relieve the pressures currently distorting our grading. Because faculty feel pressured by *Q* scores and enrollment numbers, we must make sure that our review processes reward teaching that is rigorous as well as meaningful—and that our enrollment expectations are reasonable and clear. Because faculty feel pressured by their concern for struggling students, we must make sure that these students get the support they need outside the classroom so that they can be held to the highest standards within. Because students feel pressured by applications and recruiting, we must make sure they understand what grades are actually required to achieve the goals they have set for themselves, how to make sense of and learn from the grades that disappoint them, how to channel the stress that is inextricable from their ambitions, and how to develop a mature attitude toward achievement. And because faculty wonder whether we have their backs, we must show, every day, that we do.

The OUE's 2023 *Report on Grading* catalogued the various efforts other colleges have made to alter grading practices. It was not an inspiring list. Yale tried raising awareness, circulating data about grades in the hopes that this would influence faculty practices. Grades stayed high. Cornell and Dartmouth tried adding more information, publishing median course grades on student transcripts in the hopes that this would encourage students to take more demanding courses. Only the most dedicated students did so. Finally, Princeton and Wellesley tried hard limits, capping the number of A grades that could be awarded in a single course. Within ten years, the faculty at both colleges voted to lift the caps. This history might seem to suggest that grading is an insoluble problem; what it really shows is that grading is a problem too complex to admit a single solution. Our approach bears that in mind. We are trying awareness raising, information sharing, and capping A+s all at the same time, and we are doing more than that as well. By reducing the pressures on grading, by altering the

incentives, and by deciding to act collectively to solve the collective action problem, we can restore our grading to what it was before the dramatic changes of recent years.

IV. What We Owe Our Students

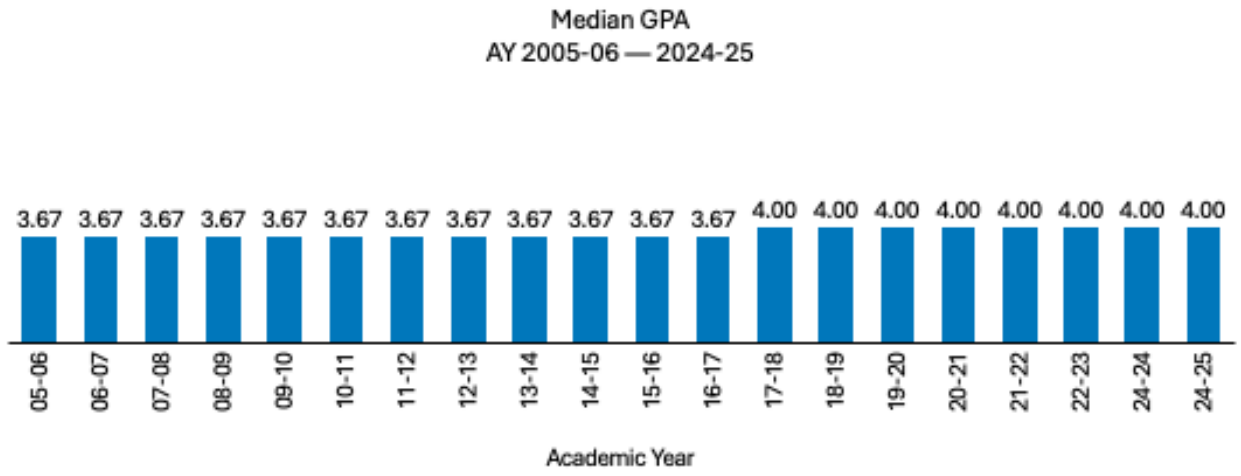
We owe our students a functioning grading system. Specifically, we owe them grades that send clear signals, that give them a good sense of their strengths and weaknesses and that communicate their areas of distinction to employers and admissions committees.

We owe our students much more than that. We owe them an education that is meaningful as well as rigorous; we owe them an education that feels "worth it" to them. That's a phrase that comes up often in conversation with students, who defined that feeling in various ways. For some, a course felt "worth it" when the faculty seemed to care about their learning. These students spoke of seminar leaders who met with them over lunch or dinner, of lecturers who lingered before and after class to talk more informally. For others, what mattered was a "tangible takeaway." Students used this phrase to describe frameworks or skills they could take from one course into another. And for another group of students, what mattered was a sense of connection between the classroom and the world. The specifics matter less, though, than the desire they reveal: students have come to us hoping for a meaningful education, and they are willing to embrace the rigor that such an education requires.

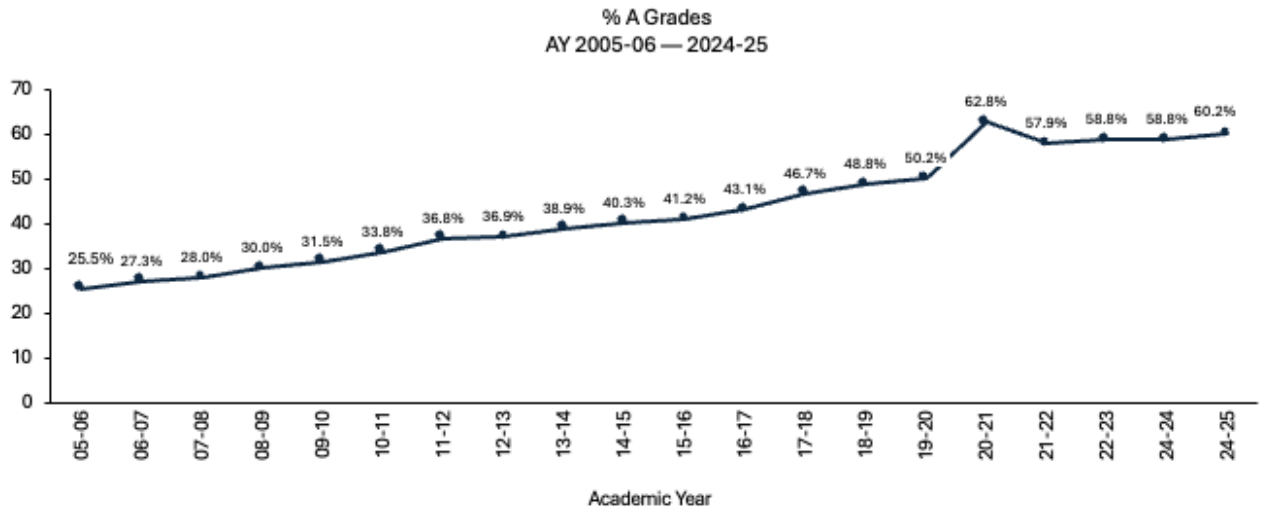
Our students are as talented and capable as they've ever been; we're more dedicated to our teaching than ever before. Together, we can reverse the drift of the past ten years and return academics to the center of undergraduate life.

Appendix

1) FAS Office of Institutional Research, “Median GPA, 2006-2025”.



2) FAS Office of Institutional Research, “Percentage of A Grades, 2006-2025”.



3) “FAS Grading Guidelines,” Harvard College Student Handbook, 2024–2025, 29-30.

Grades and Honors

GRADING SYSTEM

The FAS uses the following system of letter and non-letter grades to evaluate undergraduate student work:

LETTER GRADES

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.

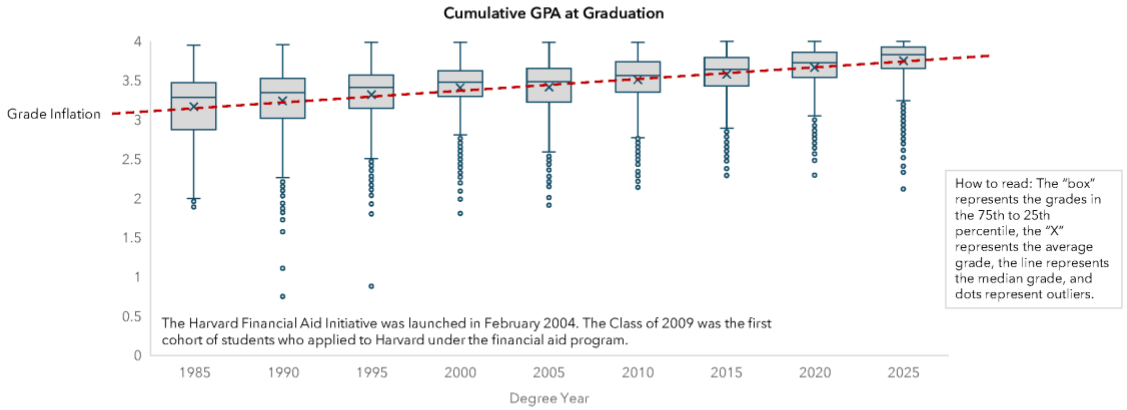
E Earned by work that is unsatisfactory and unworthy of course credit toward the degree.

4) FAS Office of Institutional Research, “Grade Compression and Grade Inflation, 1985–2025.”

Grade Inflation

Grade inflation is the increase in the average or median grade without a corresponding increase in student learning or achievement.

In 1985, the average cumulative GPA among graduating students was 3.17*. In 2025, the average GPA of graduating students increased to 3.75. Median GPA increased from 3.29 (in 1985) to 3.83 (in 2025).



	1985*	1990*	1995*	2000*	2005	2010	2015	2020	2025
Mean	3.17	3.24	3.32	3.41	3.42	3.51	3.58	3.67	3.75
Median	**3.29**	**3.35**	**3.42**	**3.48**	**3.49**	**3.56**	**3.64**	**3.73**	**3.83**
75th percentile	3.48	3.53	3.57	3.63	3.66	3.74	3.80	3.86	3.93
25th percentile	2.87	3.02	3.15	3.30	3.23	3.35	3.43	3.54	3.66

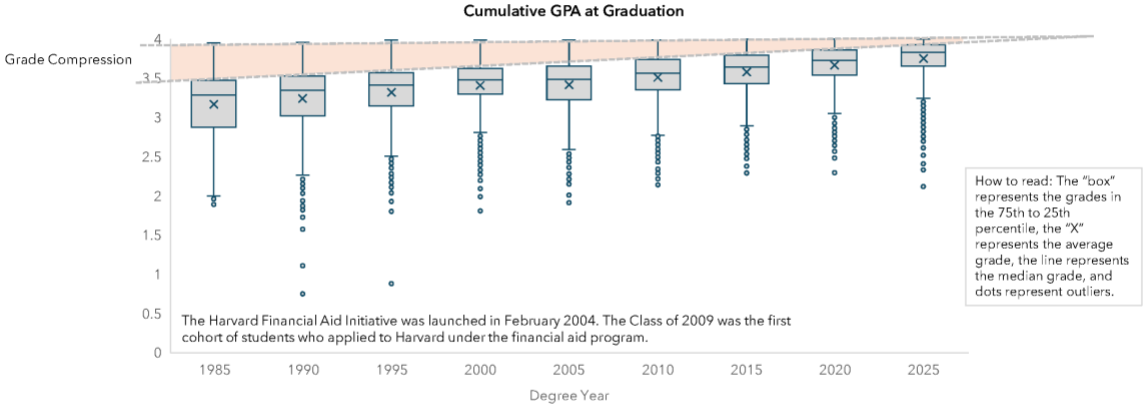
*1985 – 2004: Grades converted from 15-point scale to 4-point scale

FAS IR 2

Grade Compression

Grade compression is the narrowing of the distribution of grades awarded, especially at the high end of the scale.

Using cumulative GPA at graduation, **grade compression** can be seen in the narrowing of the range in cumulative GPA of students in the top 25% of the graduating class. In 1985, the range for the top 25% of the class was 3.48 to 3.95 (range = 0.47). In 2025, the range for the top 25% of the class narrowed to between 3.93 and 4.0 (range = 0.07).



	1985*	1990*	1995*	2000*	2005	2010	2015	2020	2025
Mean	3.17	3.24	3.32	3.41	3.42	3.51	3.58	3.67	3.75
Median	**3.29**	**3.35**	**3.42**	**3.48**	**3.49**	**3.56**	**3.64**	**3.73**	**3.83**
75th percentile	3.48	3.53	3.57	3.63	3.66	3.74	3.80	3.86	3.93
25th percentile	2.87	3.02	3.15	3.30	3.23	3.35	3.43	3.54	3.66

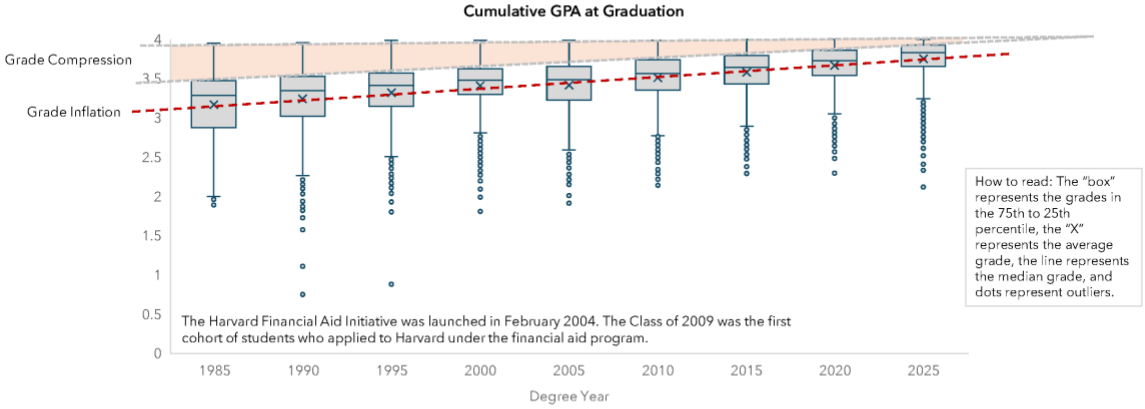
*1985 – 2004: Grades converted from 15-point scale to 4-point scale

FAS IR 1

Grade Inflation and Grade Compression

Grade compression is the narrowing of the distribution of grades awarded, especially at the high end of the scale.

Grade inflation is the increase in the average or median grade without a corresponding increase in student learning or achievement.

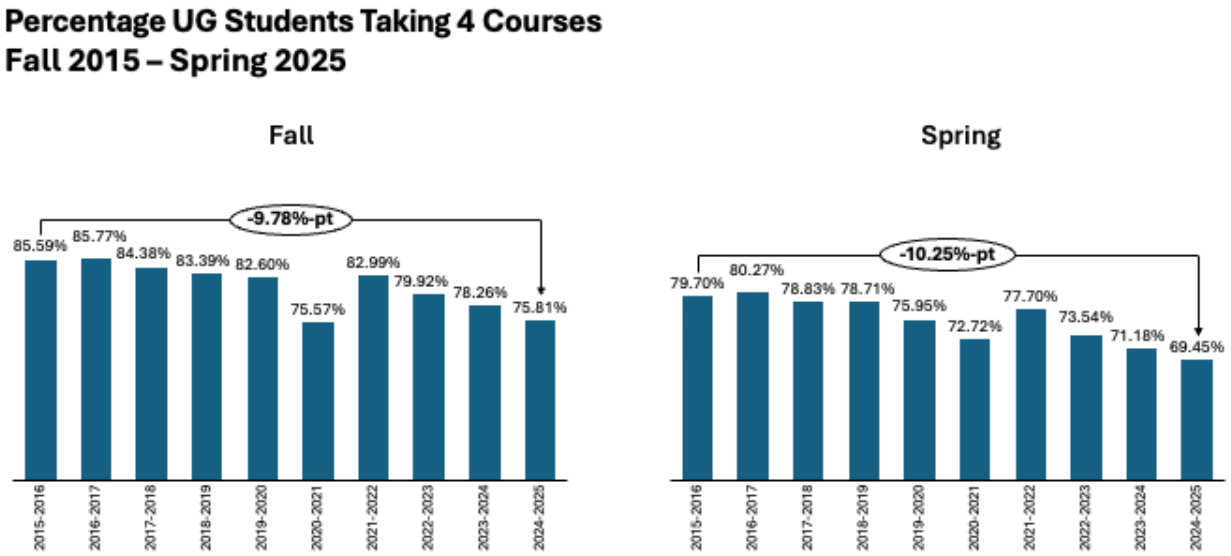


	1985*	1990*	1995*	2000*	2005	2010	2015	2020	2025
Mean	3.17	3.24	3.32	3.41	3.42	3.51	3.58	3.67	3.75
Median	**3.29**	**3.35**	**3.42**	**3.48**	**3.49**	**3.56**	**3.64**	**3.73**	**3.83**
75th percentile	3.48	3.53	3.57	3.63	3.66	3.74	3.80	3.86	3.93
25th percentile	2.87	3.02	3.15	3.30	3.23	3.35	3.43	3.54	3.66

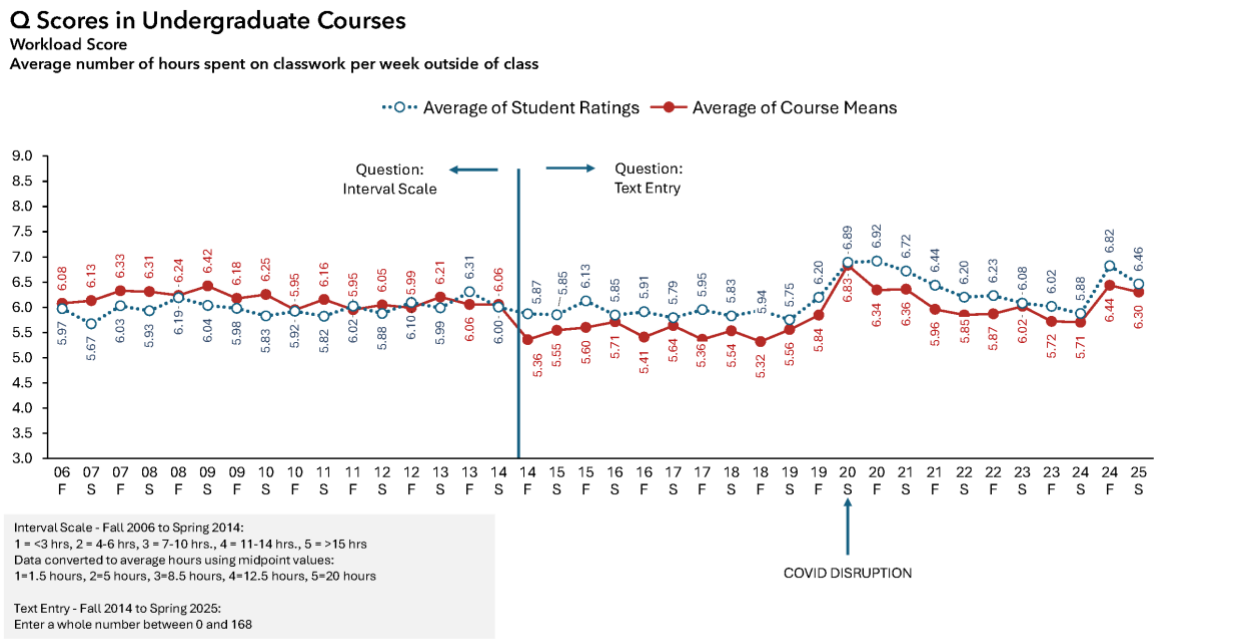
*1985 – 2004: Grades converted from 15-point scale to 4-point scale

FAS IR 3

5) FAS Office of Institutional Research, “Courses Per Semester, Fall 2015–Spring 2025.”



6) FAS Office of Institutional Research, “Workload Trends, Fall 2005–Spring 2025.”



Note: Average of course means is shown alongside the raw average of the student ratings. In the average of course means, each course is represented once in the overall average for the semester, regardless of size of course. The raw average of student ratings is the simple average of all ratings given during the semesters. In this calculation, the overall average for the semester will be skewed towards the experience of students in the larger classes.

FAS-IR

7) Lisa Mincielli and Nina Zipser, “Do Difficulty, Workload, and Grades Correlate with Instructor *Q* Scores?”, 3 March 2025.

MEMORANDUM

To: Academic Deans

From: Nina Zipser and Lisa Mincieli

Date: March 3, 2025

Subject: Do difficulty, workload, and grades correlate with instructor Q scores?

Introduction

This memo analyzes the relationship between Q “instructor scores” and three key factors: difficulty, workload, and student course grades. The dataset covers Fall 2015 to Fall 2023, beginning after the “difficulty” rating was removed from student view. Our final sample consists of 20,594 faculty-course observations, drawn from 408,974 evaluations completed by 32,545 students.[1]

We analyze these relationships using an Ordinary Least Squares (OLS) regression model. The key independent variables include:

- **Difficulty score** (1-5 scale)
- **Workload** (average hours spent per week outside of class)
- **Average student grade** in the course (0-4 scale)

We control for several additional factors known to influence instructor scores (Zipser et al., 2021):

- Course division
- Enrollment (log-transformed)
- Fraction of students taking the course as an elective
- Instructor rank (Tenured Professor, Associate Professor, Assistant Professor, Non-ladder and Visiting Faculty, Non-FAS faculty)
- Year-term
- Fraction of students taking the course pass/fail or sat/unsat

The OLS regression model is specified as follows[2]:

---

[1] For each course a faculty member has taught in the FAS, we create a “faculty-course”, which is a unique combination of semester, course catalog number and instructor ID (Zipser et al., 2021). This approach allows courses that are co-taught to be included in the dataset as discrete listings for each instructor.

[2] We also ran a version of the model without a control for average grade and fraction of students taking the course pass/fail. The same patterns and statistically significant findings persist.

Average Instructor Score = β_0 + β_1(average difficulty) + β_2(average difficulty)^2 + β_3(average hours worked) + β_4(average grade) + β_5(course division) + β_6(ln(enrollment)) + β_7(fraction elective) + β_8(instructor rank) + β_9(year-term) + β_{10}(fraction pass/fail)

Results

Table 1 (attached at the end of this memo) presents the results of the OLS regression model. We summarize our findings here (controlling for the other variables in the regression equation above):

- Difficulty essentially has no meaningful predictive relationship with instructor score, because the quadratic (i.e., curved) relationship that we estimate is almost flat in the relevant range of variation for course difficulty.
- Average grade has very small predictive power for instructor score. A one standard deviation decrease in the average course grade is associated with an estimated 0.033-point drop in instructor score (on the 1-5 Likert scale).
- Workload has no statistically significantly predictive relationship with instructor score.

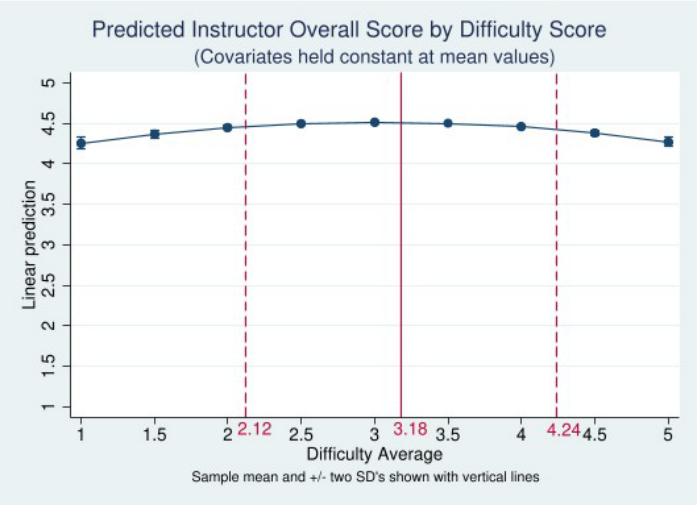
We now discuss these findings in more detail:

Difficulty

Difficulty, measured on a 1-5 scale, statistically significantly predicts instructor scores. Difficulty has a positive coefficient ($p < 0.000$), while Difficulty squared has a negative coefficient ($p < 0.000$), indicating a non-linear (curved) relationship. **However, the effect sizes are very small, so the relationship remains essentially flat within two standard deviations of the mean difficulty rating**.

- Mean difficulty: 3.18
- Standard deviation: 0.53

The graph below illustrates the correlation between difficulty and instructor score while holding other variables constant at their sample means. The solid red line represents the mean level of Difficulty, and the two dotted red lines each represent two-standard-deviation shifts, respectively, below and above the mean.



The adjusted R^2 of the model is 0.162, meaning the included covariates explain 16.2% of the variation in the average instructor score (at the course level). If we exclude average Difficulty and average Difficulty squared, the adjusted R^2 drops to 0.158, **indicating that difficulty accounts for less than half a percentage point of the variation in instructor scores.**

Workload

Workload, measured as the average number of hours students spend outside of class per week on the course, is **not statistically significant (p = 0.148)**, meaning it does not statistically significantly predict instructor scores.

- **Mean workload**: 6.57 hours per week
- **Standard deviation**: 3.5 hours

Grades

Average course grades (on a 0-4 scale)[3] statistically significantly predict instructor scores ($p < 0.000$), with a coefficient of 0.115.

- **Mean grade**: 3.80
- **Standard deviation**: 0.29 (approximately one-third of a letter grade)

[3] 0.08% of grades in our sample were coded as >4. These were recoded to 4.

A one standard deviation decrease in the average course grade is associated with an estimated 0.033-point drop in instructor score. To take an extreme example, holding all else equal, comparing two courses one of which has an average course grade that is lower by one full letter grade (approximately 3.5 standard deviations), that course's instructor score is predicted to be 0.1 points lower on the 1-5 Likert scale.

If we exclude average course grade, the adjusted R^2 decreases to 0.15, indicating that course grades account for approximately 1% of the variation in instructor scores.

Summary

While difficulty and grades statistically significantly predict instructor Q scores, the effect sizes are small. These variables explain respectively 0.5% to 1% of the variation in instructor scores. Moreover, in the relevant range of variation, difficulty has almost no material relationship with instructor score.

Workload, measured by the average number of hours students spend on a course, does not statistically significantly predict instructor Q scores.

Overall, our findings suggest that difficulty, grades, and workload are not meaningful factors predicting instructor Q scores.

Table 1. Regression results

	Coefficient	*Std. err.*	*P>t*
Average Difficulty	0.383	0.051	0.000
(Average Difficulty)^2	-0.063	0.008	0.000
Average Hours Worked	0.002	0.001	0.148
Average numeric grade	0.115	0.012	0.000
Division (Humanities ref.)			
Science	-0.183	0.009	0.000
Social Science	-0.102	0.009	0.000
Expository Writing	-0.025	0.020	0.204
General Education	-0.117	0.018	0.000
SEAS	-0.286	0.014	0.000
ln(Enrollment)	-0.068	0.003	0.000
Fraction elective	0.202	0.012	0.000
Faculty Rank (tenured ref.)			
Nonladder/Other	0.041	0.008	0.000
Assistant Professor	0.058	0.014	0.000
Associate Professor	0.067	0.019	0.000
Non-FAS	-0.043	0.010	0.000
Year-term (Fall 2015 ref.)[4]			
Spring 2016	0.010	0.017	0.572
Fall 2016	0.039	0.017	0.024
Spring 2017	0.037	0.017	0.030
Fall 2017	0.024	0.017	0.159
Spring 2018	0.069	0.017	0.000
Fall 2018	0.046	0.017	0.008
Spring 2019	0.050	0.017	0.003
Fall 2019	0.154	0.018	0.000
Fall 2020	0.218	0.018	0.000
Spring 2021	0.214	0.018	0.000
Fall 2021	0.174	0.018	0.000
Spring 2022	0.176	0.019	0.000
Fall 2022	0.210	0.018	0.000
Spring 2023	0.214	0.018	0.000
Fall 2023	0.210	0.018	0.000
Percent Pass/Fail	-0.322	0.059	0.000
Constant	3.583	0.096	0.000
N	20594		
R-Square	0.162		

[4] Spring 2020 is not included in the analysis as students did not complete the standard Q evaluation due to the Covid-19 pandemic.