The Grade Multiplier

Applying Gresham's law to grade inflation

Tiffany Mikka

Research : 10 March 2021 *Preliminary draft* : 07 April 2021

Department of Economics
University of California, Los Angeles

ABSTRACT

In search of the grade multiplier

When “A” is for average and Gresham’s law (of grade inflation) posits that “bad grades drive out good grades”

An “A” today is not worth the same as an “A” a few decades ago. Indeed, GPAs seem to be among the list of highly depreciable intangibles. No higher-education institution appears to be immune to the wave of rampantly rising grades. While nationwide numbers are indubitably rising over time, the significance of such a phenomenon may be harder to discern. With two sets of grade distribution and corresponding course evaluation data from University of California, Los Angeles, various aspects of grading patterns and responses were tested for statistical significance and analysed in the total of ten regression models to follow. The first set of data utilises Latin honours eligibility criteria for the College of Letters and Sciences (L&S) and the Henry Samueli School of Engineering and Applied Sciences (HSSEAS) comprising the years 1995-2021 and 2005-2021, respectively, in order to prove the prevalence of grade inflation in postsecondary education. In the second data set, each of the eight regression models tested a distinctive combination of variables selected from the 38 respective variables included among 106 observations. All included variables are statistically significant at less than or equal to the 5% level of significance. Results from both data sets reinforce the reasonings and incentives for grade-inflating, some of which are briefly described below. Most significantly, when interpreted in the context of the signalling game (*see subsection*: **Signalling askew**), findings suggest attenuation in the signalling value of grades through (1) illustrating the theoretical convergence of all grades towards the upper limit of the current 4.0 GPA cap and (2) strongly supporting the incentive from instructors to grade-inflate. This attenuation in signalling value aptly reflects the notion of Gresham’s law[1], where the bad drives out the good: in this case, bad grades accomplish this by replacing the good through devaluation. To be sure, this preliminary analysis alone can neither prove nor disprove the existence of grade signal attenuation. Nevertheless, as this research intends, some critical underlying elements are identified to be significantly influential in the context of grade inflation and signalling value, and a prevalent yet often-overlooked national phenomenon potentially affecting 17 million undergraduate students[2] is forced into the spotlight, front and center.

Note: The subsequent data and preliminary analyses are intended to supplement the corresponding main article *The Grade Multiplier, applying Gresham’s law to grade inflation* (April 2021).

Acknowledgements

As a right-brained individual with absolutely no coding or statistics background, I never would have believed that I would one day fall in love with running regressions on random data - not out of necessity to turn in an assignment or for fear of failing a course - but purely out of curiosity and captivation.

A sincere thank you to Dr. Randall R. Rojas for introducing me to statistics, regressions, and programming, and inspiring me to further explore the wondrous world of data analysis. Spellbound as I am now, I will not ever forget the week 1 struggles - how stressed, fearful, and dreadful I felt imagining how I would deal with programming for the next 10 weeks. I truly am absolutely grateful for all the valuable insight, skills, knowledge, and tools I have taken away from a mere 10 weeks - and above all, for the encouragement and support in developing my newfound interest in a field that I was, frankly, previously utterly terrified by.

[1]For an interesting bit of inspiration on this principle (largely applied in the context of investing and finance, where “bad money drives out good money”), see “Gresham’s law: Why Bad Drives Out Good As Times Passes”. *Farnam Street Media*, December 2009.

[2]Based on a projection by the National Center for Education Statistics for total undergraduate student enrollment in 2029: “Undergraduate Enrollment.” *National Center for Education Statistics (NCES)*, May 2020.

A special thank you to *Daily Bruin* for providing the Bruinwalk grade distribution data - this project could not have come together without both halves of the story (*i.e.*, grade distributions and course evaluations).

The biggest thank you especially to Jonathan Rieck for all the insightful input and analysis ideas, organising the (originally) very messy mélange of raw data files, and permitting me to run off with and write an obnoxiously long report on your idea (of a potential grade multiplier). Certainly, no such concept could have sprang up in any other conversation. And as I have a tendency to not follow through on projects I (impulsively) pick up, thank you above all for shepherding this little one along its way.

Introduction

A popular character in the Keynesian story is the government multiplier (g), denoted by $\Delta Y = g\Delta G$, where g is given by $\frac{1}{1-MPC}$. Among some of the siblings are t (the tax multiplier) and m (the open economy multiplier), and perhaps new member may be joining the family: the grade multiplier.

In the context of the educational market, the grade multiplier would represent a similar phenomenon: an increase in average grade point by a point increase triggering an increase in the overall economic output of the increase multiplied by a factor. The translation may be interpreted as increasing grades indicate better jobs obtained (where high-skilled jobs imply market efficiency and thus economic output expansion). However, it turns out that the grade multiplier may, in actuality, be negative: grade increases (or "inflation") may be dampening the job market with worse jobs and, au contraire, contractionary.

To follow are the steps set out in search of such a multiplier.

Background

As with monetary currency, grades have been on the rise for decades. The widespread phenomenon plagues the vast majority of schools, especially institutions with higher selectivity. Among the most frequently cited and chastised for rampant grade inflation is Harvard University, where 91% of its graduating class receives "honours." In inquiring about Harvard College's grade distribution, professor Harvey Mansfield said to the Dean of Undergraduate Education: "A little bird has told me that the most frequently given grade at Harvard College right now is an A-."

The Dean's apparent correction:

> "The median grade in Harvard College is indeed an A-. The most frequently awarded grade in Harvard College is actually a straight A."

But as illustrated in the corresponding literary analysis and economic reconstruction of the grade inflation phenomenon, not all inflation is created equally: monetary currencies can, theoretically, rise indefinitely, while grades, au contraire, are capped at a ceiling (equating to the maximum GPA of 4.0). The less-common concept of grade compression will thus be introduced in subsequent analyses to further emphasise this distinction.

To date, a vast assortment of reasonings exist in current literature to explain the prevalence of grade inflation at the undergraduate level[3].

[3] *See subsection*: The inflation incentive, from the main article *The grade multiplier, applying Gresham's law to grade inflation.*

Historically, grade inflation is said to have arisen from the Vietnam era[4], where higher grades "shielded" a student from being drafted into the Vietnam War[5]. Teachers were thus said to have given students the minimum B mark needed to spare them from the draft.

However, as the war faded into distant memory, grades nevertheless continued to spring upwards. This second era of major grade inflation, spanning from around 1983 to 2013[6], reflects the rise of consumerism[7] and, in this context, the increasing commercialisation of American postsecondary education. Higher education is becoming more of a business than ever, and along with price tags come expectations - notably, the guarantee of landing a high-paying job upon graduation. Students seek high grades that, presumably, will convert to securing a high-paying job, while schools search for students with the most potential in order to boost their own reputations. Logically, the solution is to grade-inflate: both the student and the school obtain their individual objectives, and the two-way condition is satisfied.

On a similar note, financial pressures may function as another influential factor: financial aid scholarships and tax credits[8] oftentimes require a minimum GPA for consideration. Understandably, instructors are reluctant in denying such students the B mark needed for a tax break or tuition deduction. In the state of Georgia, for example, students must maintain a minimum B average in order to qualify for a merit-aid scholarship[9]. Failure to do so and losing financial aid may jeopardise their enrollment status.

An additional element influencing grade inflation relates to the tenure status of course instructors. Typically, most university instructors seek to obtain tenure, which promises better pay and job stability, among other benefits. The process is largely based on research output, but student evaluations are also factored in, as an indicator of faculty instruction performance. This creates a further interesting phenomenon: while commonly poor students hoping for a good grade are envisioned to be at the mercy of mean professors, the converse is hardly imaginable - poor professors trying to obtain tenure being at the mercy of picky students' evaluation scores. But this clearly illustrates why instructors - especially untenured instructors - would have a substantial incentive to grade-inflate and thus earn higher evaluation scores.

The basic plot to follow highlights the grade inflation phenomenon at the University of California, Los Angeles, from 1927 to 2015[10]. Both lower and upper division courses are included from the spring term of each included year.

```
## Warning: not enough colors. Will repeat.
```

[4]This is a commonly-cited explanation currently in circulation. See the plot to follow (at the end of this section) for a visual representation of this reasoning.

[5]Jacobs, P. (2013). *Many College Professors Started Using Grade Inflation To Protect Bad Students From Being Drafted Into The Vietnam War*. *Business Insider*.

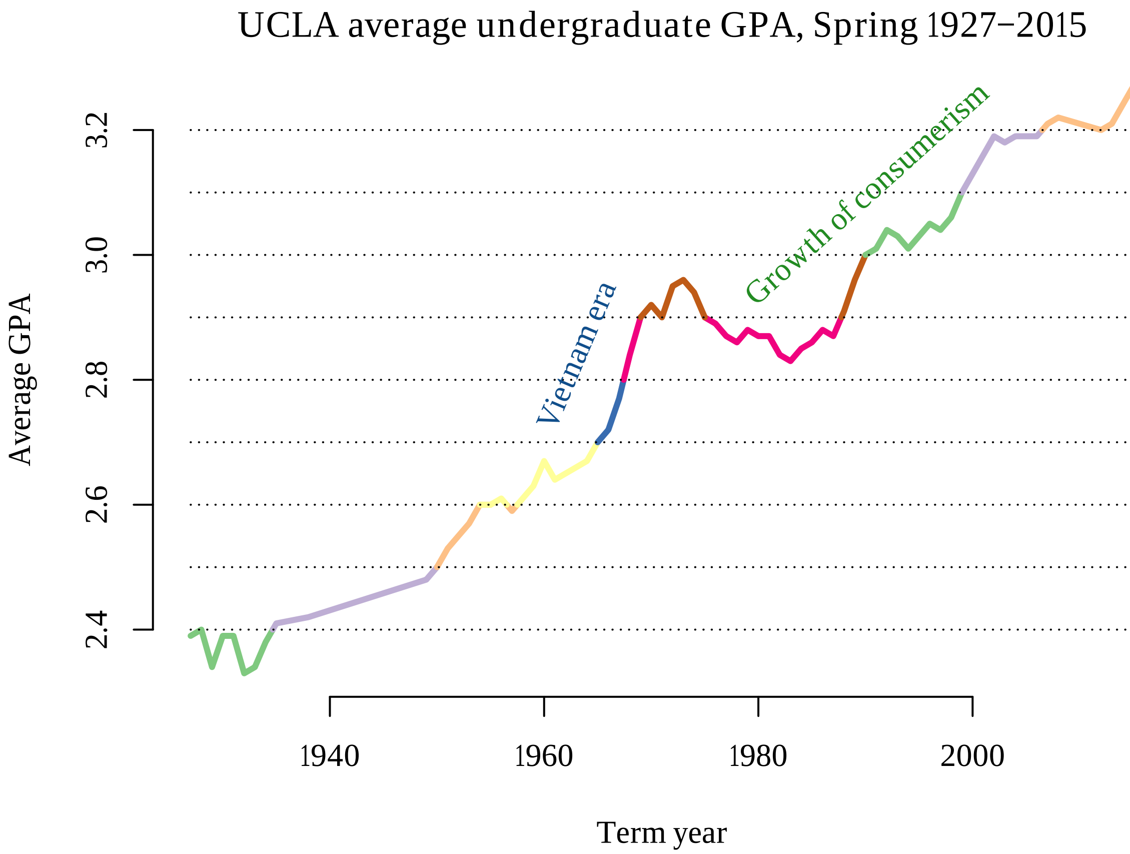
[6]See the following plot for an illustration of this widely-proposed pattern in existing literature on causes of grade inflation.

[7]For a more in-depth discussion on approaching grade inflation in the context of consumerism, see

[8]Reischauer, L. E. & Gladieux, R. D. (1996). *Higher Tuition, More Grade Inflation*. Brookings.

[9]Mathies, C., Bauer, K. W., and Allen., M. (2005). *Thoughts on Grade Inflation*. University of Georgia.

[10]Data extracted from Rojstaczer, S. *Grade inflation*. Stuart Rojstaczer is a former Duke professor who compiled grade data from a wide range of schools spanning across multiple decades. The full list of schools included in this study may be found on the *main page*.



Note: Each colour interval denotes a 2.5% (0.1-point) increase in grade point average (GPA).

I. Grades on the Rise

Variable Selection

Latin Honours (*cum laude*, *magna cum laude*, and *summa cum laude*) are awarded to the top 20%, 10%, and 5% of a graduating class, respectively. Because percentile ranks generally remain a stable measurement of student achievement over time, the eligibility criteria would be a fairly reliable indicator of grade inflation by functioning as a control for student achievement. The independent variable in each model is thus given by the set of minimum GPA thresholds in order to be eligible for each respective recognition of cum laude, magna cum laude, and summa cum laude.

Data

Utilising two sets of data extracted from (1) the College of Letters and Sciences (L&S) and (2) Henry Samueli School of Engineering and Applied Sciences (HSSEAS) at University of California, Los Angeles, the following simple linear regression (applicable to both data sets) was built and subsequently analysed:

$$LATHON = \lambda_1 + \lambda_2 GRADYEAR + e, \tag{1}$$

where $LATHON$ is substituted by the respective recognitions of *cum laude*, *magna cum laude*, and *summa cum laude*.

Figures

Three separate linear regressions, all of the above type, were constructed for each set of data, shown below in tables and as a plot.

Note: The data set for HSSEAS would fall under a STEM classification. However, the L&S data set does not constitute purely non-STEM majors; hence, the two sets of data are not labelled as "STEM" and "non-STEM."

[i] College of Letters and Sciences The College of Letters and Sciences includes a mélange of majors, including arts, STEM, and interdisciplinary majors such as:

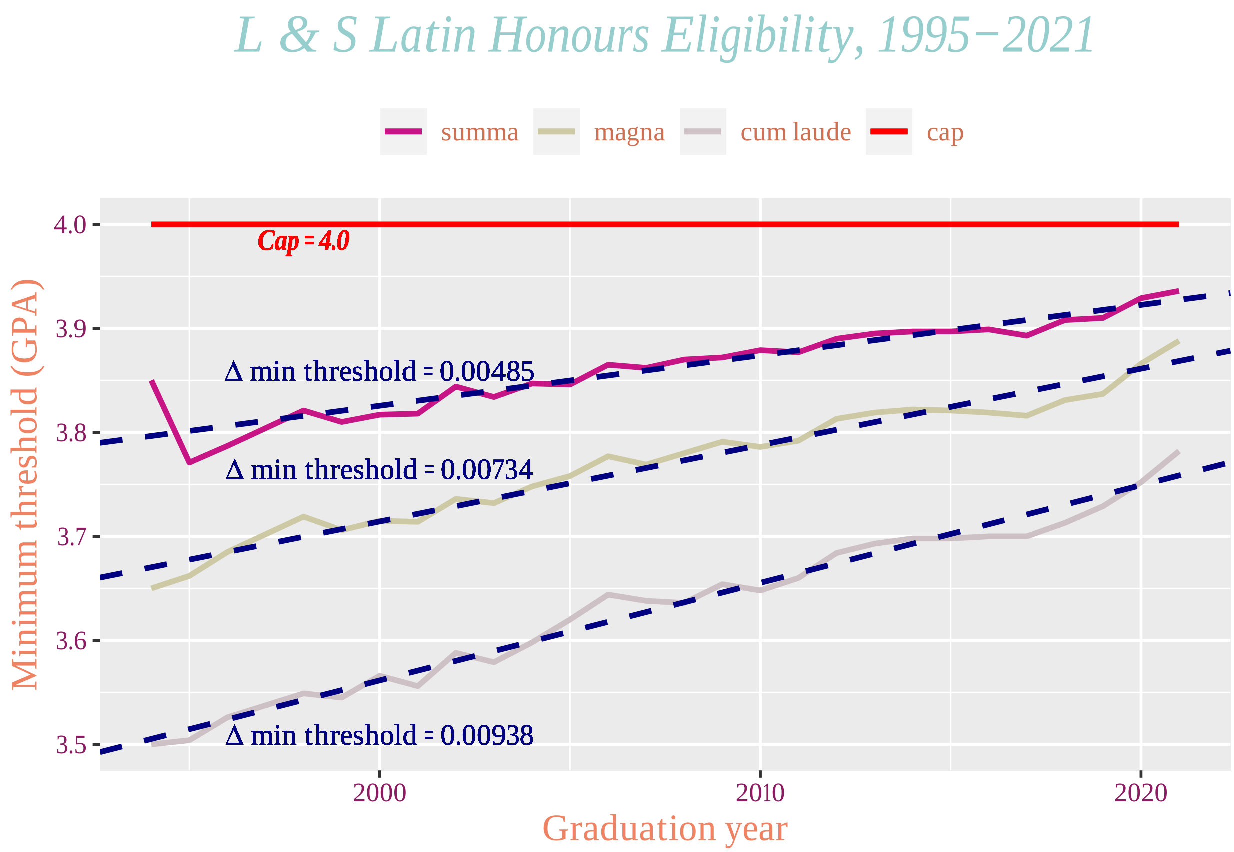
- Anthropology (B.A.)
- Astrophysics (B.S.)
- Biology (B.S.)
- Business Economics (B.A.)
- Comparative Literature (B.A.)
- Computer Science
- Economics (B.A.)
- English (B.A.)
- Geology (B.S.)
- International Development Studies (B.A.)
- Linguistics (B.A.)
- Mathematics (B.S.)
- Neuroscience (B.S.)
- Political Science (B.A.)
- Psychology (B.A.)
- Statistics (B.S.)

A full list of departments and programs may be found on the College's ***Academics*** page.

Table 2: Letters and Sciences Latin honours eligibility, 1995-2021

	Dependent variable:		
	cum laude	magna cum laude	summa cum laude
	(1)	(2)	(3)
Graduation year	0.009***	0.007***	0.005***
	(0.0003)	(0.0003)	(0.0004)
Constant	−15.205***	−10.974***	−5.870***
	(0.575)	(0.583)	(0.711)
Observations	27	27	27
R^2	0.977	0.962	0.882
Adjusted R^2	0.976	0.961	0.878
Residual Std. Error (df = 25)	0.012	0.012	0.015
F Statistic (df = 1; 25)	1,072.564***	639.143***	187.419***
Note:			*p<0.1; **p<0.05; ***p<0.01

Note: The academic year 1996-1997 ($GRADYEAR = 1997$) was dropped due to missing observations.



Note: The academic year 1996-1997 ($GRADYEAR = 1997$) was dropped due to missing observations.

Results

Of the three sets of data, all were statistically significant, as follows:

$$\underset{(se)}{LAUDE} = \underset{(0.575)^{***}}{\gamma_1} + \underset{(0.003)^{***}}{\gamma_2} \; GRADYEAR; \tag{2}$$

$$\underset{(se)}{MAGNA} = \underset{(0.583)^{***}}{\beta_1} + \underset{(0.003)^{***}}{\beta_2} \; GRADYEAR \tag{3}$$

$$\underset{(se)}{SUMMA} = \underset{(0.711)^{***}}{\alpha_1} + \underset{(0.004)^{***}}{\alpha_2} \; GRADYEAR \tag{4}$$

Refer to the statistical summary tables.

Results: The minimum threshold to receive *cum laude* increased by the most, which an increase in grade point of approximately 0.00938 for every additional year. The minimum threshold for *magna cum laude* had the next greatest overall change of approximately a 0.00734 grade point increase for each additional year. The minimum threshold for *summa cum laude* increased by the least, with an increase in grade point of 0.00485 per additional year. Since all 3 of the 3 honours demonstrated a statistically significant increase in threshold value, there does appear to be significant grade inflation per the regression models built for these data sets.

Analysis: The change in minimum threshold for each of the Latin honours, all significant at the 0.1% level, precisely illustrate the case of grade compression: the top scores at capped at a ceiling threshold (of 4.0 in the GPA context), while the lower scores are allowed to float upwards. Theoretically, the minimum threshold for *cum laude* could continue to increase until it converges with *summa cum laude*, which has the highest minimum threshold, since minimum threshold for *summa cum laude* cannot continue to increase past the 4.0 ceiling. In other words, *summa cum laude* is restricted from further increasing by the GPA cap, while *cum laude* and *magna cum laude* still have room to continue increasing. The GPA cap restricting the increase of *summa cum laude* but not as much for *magna cum laude* and *cum laude* may be seen from the observation that *cum laude* is increasing at the greatest rate, while *summa cum laude* is increasing at the least rate, and *magna cum laude* falls somewhere in between (though expectedly closer to the rate of *cum laude*).

These Latin honour eligibility thresholds may be applied to the broader context of grades. For simplification, we can assume *summa cum laude* to represent the proportion of A grades, *magna cum laude* to be the share of B grades, and *cum laude* to indicate C grades. Since the A grades are capped at a ceiling (of, clearly, an A), the B and C grades theoretically can catch up to the A grades if grades are permitted to continuously rise. The gap between C and A grades will increasingly decrease, until C grades eventually converge with A grades.

Mathematically, we can say

$$\lim_{GRADYEAR \to \infty} LATHON = 4.0, \text{where } LATHON = SUMMA, MAGNA, LAUDE$$

$$\therefore \lim_{GRADYEAR \to \infty} SUMMA = MAGNA = LAUDE = 4.0$$

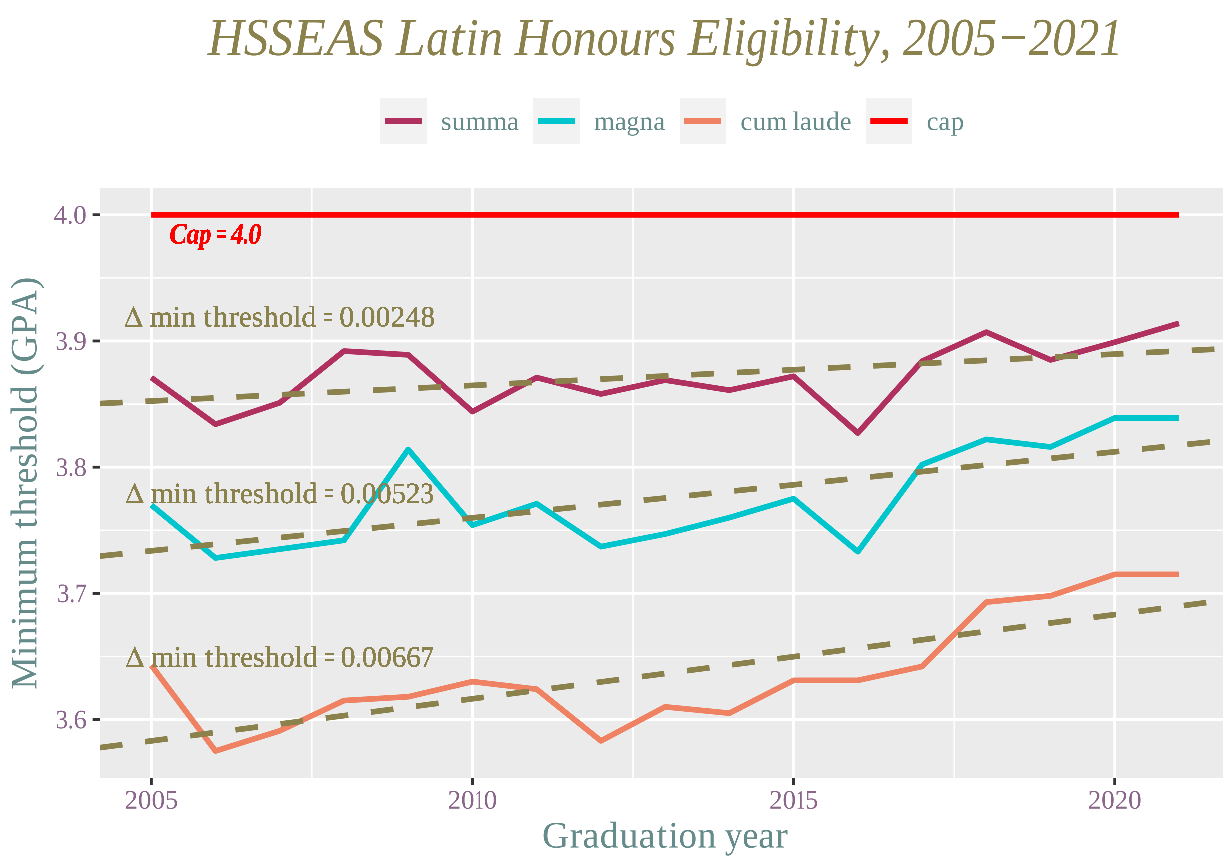
[ii] Henry Samueli School of Engineering and Applied Sciences HSSEAS contains STEM majors and thus tends to have lower average GPAs. Majors include:

- Aerospace engineering
- Bioengineering
- Chemical engineering
- Computer science
- Electrical engineering
- Environmental engineering
- Materials engineering
- Mechanical engineering

A comprehensive list of current undergraduate majors may be found on the school's ***Research and Admissions*** website.

Table 3: HSSEAS Latin honours eligibility, 2005-2021

	Dependent variable:		
	cum laude	magna cum laude	summa cum laude
	(1)	(2)	(3)
Graduation year	0.007***	0.005***	0.002**
	(0.001)	(0.001)	(0.001)
Constant	−9.798***	−6.758**	−1.121
	(2.896)	(2.906)	(2.188)
Observations	17	17	17
R^2	0.589	0.467	0.258
Adjusted R^2	0.562	0.431	0.208
Residual Std. Error (df = 15)	0.029	0.029	0.022
F Statistic (df = 1; 15)	21.514***	13.141***	5.207**
Note:			*p<0.1; **p<0.05; ***p<0.01



Results

Of the three sets of data, all were statistically significant, as follows:

$$\underset{(se)}{LAUDE} = \underset{(2.896)^{***}}{\Gamma_1} + \underset{(0.001)^{***}}{\Gamma_2} GRADYEAR; \tag{5}$$

$$\underset{(se)}{MAGNA} = \underset{(2.906)^{**}}{B_1} + \underset{(0.001)^{***}}{B_2} GRADYEAR \tag{6}$$

$$\underset{(se)}{SUMMA} = \underset{(2.188)}{A_1} + \underset{(0.001)^{**}}{A_2} GRADYEAR \tag{7}$$

Refer to the statistical summary tables.

The minimum GPA threshold to earn all three honours has increased substantially over a span of 17 years. The parameter C_2 implies that the minimum threshold for *cum laude* will increase by a grade point of approximately 0.00667 per year. Likewise, the parameter B_2, statistically significant at the 0.1% level, indicates an approximate 0.00523 grade point increase in the minimum threshold for *magna cum laude* with every additional year. The parameter of A_1 for *summa cum laude* yields an approximate 0.00248 grade point increase in average minimum GPA with an additional year.

Analysis

The data already clearly depicts the increasing trend in the honours eligibility criteria; what is of interest, instead, is the significance of this increase. In other words, an insignificant increase in the minimum GPA threshold to earn a Latin honour does not indicate an alarming prevalence of rampantly rising grades.

However, since all 3 of the 3 honours demonstrated a statistically significant increase in threshold value, there does appear to be significant grade inflation per the regression models built for these data sets.

Furthermore, the HSSEAS data depicts the same trend as with the L&S data in the preceding plot: the phenomenon of grade compression. As previously demonstrated, the lower eligibility thresholds of *cum laude* and *magna cum laude* are increasing at a greater rate than *summa cum laude*, which appears to be approaching the maximum GPA limit of 4.0. The same mathematical equation denoting convergence by all three honours to the limit of 4.0 is similarly applicable in this case.

II. Grade inflation at the Undergraduate Level

In order to further examine grade inflation at the undergraduate level, the 8 subsequent (simple and multiple) regressions were built and analysed. Areas of interest comprise correlations between average course grade and overall rating of the course; various feature-specific ratings and the overall course rating; the easiness of a course and average course grade; the tenure status of course instructors and percentage of A grades given (*see subsection:* **The inflation incentive**); the percentage of A grades given and overall course rating; the tenure status of course instructors and total pay; and various feature-specific ratings and the percentage of A grades given.

While the overall hypothesis aims to test the significance of grade inflation at the undergraduate level, each of the following regressions tests an individual aspect of the phenomenon using a specific subset of variables included in the data set (*i.e.*, zooms in on a particular feature that may be indicative of the larger trend of grade inflation). Hence, different combinations of selected variables and separate hypotheses are included for each of the respective regressions.

Data

All subsequent regression models were built using data from Bruinwalk[11], which receives course evaluation and grade distribution data[12] from the University of California, Los Angeles Registrar's Office. The data set comprises 106 undergraduate-level courses across most departments, spanning across the years around 2017.

Variable Selection

A total of 38 variables[13] are included in the data file, among which are used (separately and jointly) in the following regression models. The variables may be categorised into 3 main groups, of which we are interested in determining the interactions and correlation between:

[1] Student achievement

- *GPA*: Average course grade, measured by a 4.0-scale grade point average;
- *Percentage As*: The percentage of students who received an A grade in the course;
- *Percentage P/NP*: The percentage of students who took the course for a Pass/No pass grade, which tends to indicate perceived difficulty. Since a P/NP course grade is not reflected on the transcript and calculated as part of overall GPA, students generally only opt to take a P/NP grade when they anticipate the course to be difficult.

[2] Course ratings

- *Overall rating*: overall course rating on a scale of 1-5;
- *Easiness rating*: a numerical rating from 1-5 indicating the leniency of a course;
- *Helpfulness rating*: a numerical rating from 1-5 indicating how helpful a course instructor was;
- *Workload rating*: a numerical rating from 1-5 indicating how much work a course was perceived to necessitate;
- *Clarity rating*: a numerical rating from 1-5 indicating how clear a course instructor was.

[3] Course characteristics

- *STEM*: An indicator variable identifying if a course is considered a STEM course;
- *Lecturer*: An indicator variable determining if a course instructor has the title of "Lecturer" and does not hold tenure status;
- *Professor*: An indicator variable determining if a course instructor has the title of "Professor" and holds tenure status;
- *Total pay*: total annual pay in dollars to the course instructor, comprising annual base pay plus other pay, less benefits.

Additional variable selection and interpretation details are included for each corresponding regression analysis.

[11] **Bruinwalk** is a course-rating platform for UCLA students managed by *Daily Bruin*. Students are able to give course ratings on a scale of 1-5 for the following course characteristics: easiness, workload, clarity, helpfulness, in addition to an overall rating.

[12] Grade distribution data is obtained for every undergraduate and graduate level course following the conclusion of each quarter by the Registrar's Office and provided to *Daily Bruin* for use on Bruinwalk.

[13] 10 of the 38 variables were added to the original data set: *lecturer*, *assoc_prof*, *prof*, *total_pay*, *upper_div*, *course_prof*, *GPA*, *percA*, *percPNP*, and *south_campus* (denoted as *STEM* for purposes of the following analyses). The variables *assoc_prof* and *upper_div* were omitted due to statistical insignificance in all models tested. The variable *course_prof* is an identification key and thus excluded in all models and analyses.

Findings

8 regression models of both simple linear and multiple regression types were constructed in order to test the hypothesis of grade inflation occurring significantly at the undergraduate level. Individual regression statistics and interpretations are summarised below in statistical summary tables and respective individual analyses for clearer organisational structure.

Model 1: Average course grade and overall rating A primary area of interest lies in the correlation between the average grade received for a particular course and the subsequent average course evaluation score. Predictably, students expecting higher grades in a course would likelier give higher ratings for the course than students anticipating lower grades. The first simple linear regression model below thus tests for this most basic effect of average course grades on average overall course evaluation scores:

$$\ln(OVERALL) = \alpha_1 + \alpha_2 GPA + e, \tag{8}$$

where $OVERALL$ represents the overall course rating on a scale of 1-5 and GPA is the average course grade on a 4.0 GPA scale.

Per the prediction that students anticipating higher grades in a course are more inclined to give higher course ratings, the following hypothesis was tested to indicate the significance of the GPA parameter:

$$H_0 : \alpha_2 = 0 \qquad\qquad H_1 : \alpha_2 \neq 0$$

Table 4: Average course grade and overall rating

	Dependent variable:
	Logged overall course rating
Average course grade (GPA)	0.157^{**}
	(0.064)
Constant	0.777^{***}
	(0.213)
Observations	106
R^2	0.055
Adjusted R^2	0.046
Residual Std. Error	0.223 (df = 104)
F Statistic	6.088^{**} (df = 1; 104)
Note:	*p<0.1; **p<0.05; ***p<0.01

Results: The subsequent statistical summary output indicates that GPA is significant at the 1% level; we reject the null hypothesis that average course grades have no impact on the average overall course rating.

$$t\text{-statistic} \approx 2.467 > t\text{-critical}_{0.975,104} \approx 1.983 \Rightarrow \textit{Reject}$$

Interpretation: A one-point increase in average course grade leads to approximately a 15.7% increase in average overall course rating.

Analysis: Students who expect to earn high grades in a course will very likely rate the course higher overall. Note that course evaluations are generally completed before final grades are submitted; thus, it is also possible that students incorrectly perceive their final grades in the course. Nevertheless, most students typically seem to have a reasonable estimate of their progress in the course based upon previous exam and assignment scores (*e.g.*, midterm exam results).

Model 2: Average course grade and overall rating, STEM vs. non-STEM majors STEM and non-STEM courses are oftentimes perceived to differ in difficulty and thus may perhaps demonstrate differing levels of grade inflation. This second model builds upon the first model testing solely the effect of average course grades on average overall course ratings to account for any potential variations between STEM and non-STEM courses. In this regression, the main interest is to determine the significance of the difference in difficulty of STEM and non-STEM courses, if applicable. This multiple log-linear regression may be represented by

$$\ln(OVERALL) = \beta_1 + \beta_2 GPA + \delta_1 STEM + e, \tag{9}$$

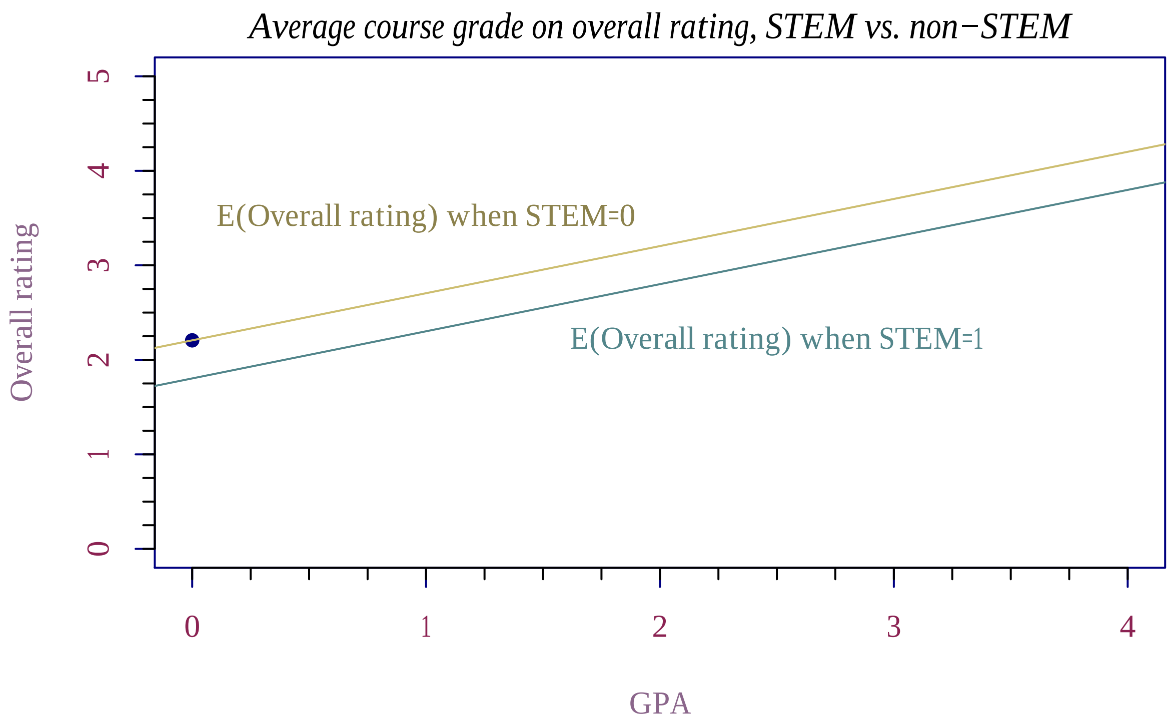
where $OVERALL$ represents the overall course rating on a scale of 1-5, GPA is the average course grade on a 4.0 GPA scale, and $STEM$ is an indicator variable denoting the presence of a STEM course.

The following hypothesis to test the overall significance of the parameters was set up as:

$$H_0 : \beta_k = 0 \text{ and } \delta_1 = 0, \text{ where } k = 1, 2 \qquad H_1 : \; \geq \text{ one } \beta_k \neq 0 \text{ and/or } \delta_1 \neq 0$$

Table 5: STEM average course grade and overall rating

	Dependent variable:
	Logged overall course rating
Average course grade (GPA)	0.153**
	(0.062)
STEM (indicator)	−0.133***
	(0.047)
Constant	0.829***
	(0.207)
Observations	106
R^2	0.123
Adjusted R^2	0.106
Residual Std. Error	0.216 (df = 103)
F Statistic	7.228*** (df = 2; 103)
Note:	*p<0.1; **p<0.05; ***p<0.01



Note: The above plot depicts unlogged overall rating values for the dependent variable (i.e., a simple linear regression) for simpler interpretation of plotted values. The log transformation used in the log-linear regression for analysis purposes contains slightly higher significance.

Results:

$$F\text{-statistic} \approx 7.228 > F\text{-critical}_{2,103} \approx 3.08 \Rightarrow Reject$$

∴ At least one of the stated parameters is statistically significant when tested jointly.

Interpretation: On average, undergraduate-level STEM courses are overall rated 13.298% worse than non-STEM courses on a numerical rating scale of 1-5 (based upon the log-linear regression). The $STEM$ parameter $\delta_1 \approx$ -0.13298, significant at the 0.1% level, gives the difference between STEM and non-STEM course overall ratings.

Analysis: There is a clear and statistically significant differentiation between STEM and non-STEM courses. As previously determined that higher course grades lead to higher overall course ratings, the converse appears to be proven in this regression analysis: lower course grades yield lower overall course ratings.

Underlying this significant differentiation is a problem of greater magnitude: the disincentivisation of students in pursuing STEM degrees[14] due to the extra burden in work and lower return in terms of GPA.

Model 3: Course grade and feature ratings on overall course rating To account for potential side effects other than average course grade, the following related variables were introduced in the next regression model: the helpfulness rating, the workload rating of STEM courses, and the clarity rating of STEM courses.

[14]This analysis is concurred in a the first part of a 4-part series on grade inflation, based upon a paper by Tom Lindsay. See Lindsay, T. (2019). *Grade Inflation in U.S. Higher Education—We Have A Problem, Part 1 of 4*. *Forbes.

This multiple log-linear regression may be modelled by

$$\ln(OVERALL) = \phi_1 + \phi_2 GPA + \delta_1 STEM + \phi_3 HLPF + \delta_2(STEM \times WKLD) + \delta_3(STEM \times CLRT) + e, \tag{10}$$

where $OVERALL$ represents the overall course rating, GPA is the average course grade on a 4.0 GPA scale, and $STEM$ is an indicator variable denoting the presence of a STEM course, $HLPF$ indicates the helpfulness rating for the instructor of the course, $WKLD$ reflects how much work the course consisted of, and $CLRT$ gives how clear the instructor was in the course. All ratings were given on a scale of 1-5.

The following hypothesis was constructed for an overall significance F-test:

$$H_0 : \phi_k = 0, \delta_k = 0, \text{ where } k = 1, 2, 3 \qquad\qquad H_1 : \;\geq\; \text{one } \phi_k \neq 0 \text{ and/or } \delta_k \neq 0$$

Results are shown in the following statistical summary table.

Table 6: STEM course features and overall rating

	Dependent variable:
	Logged overall course rating
Average course grade	0.021
	(0.037)
STEM (indicator)	-0.532^{***}
	(0.120)
Helpfulness	0.190^{***}
	(0.019)
STEM workload	0.012
	(0.048)
STEM clarity	0.133^{***}
	(0.035)
Constant	0.534^{***}
	(0.132)
Observations	106
R^2	0.709
Adjusted R^2	0.694
Residual Std. Error	0.126 (df = 100)
F Statistic	48.723^{***} (df = 5; 100)
Note:	*p<0.1; **p<0.05; ***p<0.01

Results: The parameters are jointly significant per the overall significance F-test results:

$$F\text{-statistic} \approx 48.723 > F\text{-critical}_{5,101} \approx 2.31 \Rightarrow \textit{Reject}$$

Interpretation: STEM courses still yield a lower average course grade, by approximately 50.3%. Helpfulness and STEM course clarity ratings increase average course grades by approximately 19.0% and 13.3%, respectively. Interestingly, a STEM course with greater workload also increases average course grades, by approximately 1.2%.

Analysis: Overall, STEM courses have lower average course grades compared to non-STEM courses. Factors such as how helpful the instructor is and how clear instruction is for particular course seems to positively influence average course grades, as expected.

It is interesting to note, however, that STEM courses with higher workload appear to have higher average course grades as well per the regression output. The increase is fairly small (approximately 1.2%) compared to the other predictors in the model, but nevertheless highlights an interesting point: perhaps more difficult courses with greater workloads necessitate more studying time, which converts to higher course grades. Current literature relating studying time with student achievement (as measured by indexes such as GPA) contend to the prediction that higher studying time positively correlates with higher course performance.

Model 4: Course feature ratings on average course grade Ratings for specific course features may give valuable insight into the course selection process for students, since students are able to base their decisions on previous ratings for the same course taught by the same instructor. Intuitively, one would assume that a student would be inclined to take a course that will maximise their overall GPA (just as any agent in our world of economic theory seeks to maximise individual utility, with grades being the utility in this case).

In order to determine the correlation between various course feature ratings and average student course grades, the following selection of variables were included in the multiple linear regression

$$GPA = \xi_1 + \xi_2 WKLD + \xi_3(HLPF \times EASN) + \xi_4 EASN + \xi_5(EASN \times percPNP) + e, \qquad (11)$$

where GPA is the average course grade on a 4.0 GPA scale, $WKLD$ reflects how much work the course consisted of, $HLPF$ indicates the helpfulness rating for the instructor of the course, and $EASN$ gives how lenient the course and/or professor is perceived to be. All ratings were given on a scale of 1-5.

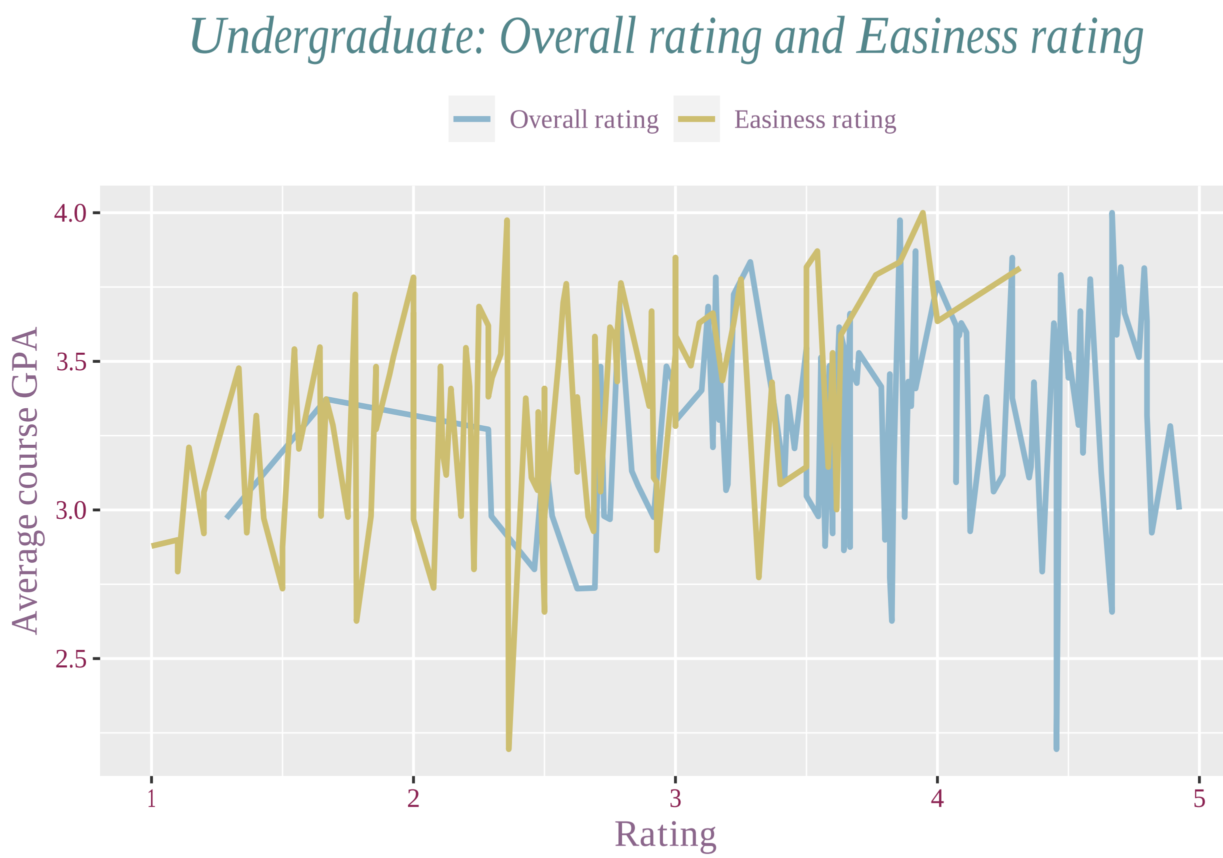
The following hypothesis was tested for overall significance:

$$H_0 : \xi_k = 0, \text{ where } k = 1, 2, 3, 4, 5 \qquad\qquad H_1 : \ \geq \text{ one } \xi_k \neq 0$$

Results are given in the following statistical summary table.

Table 7: Course feature ratings and average course grade

	Dependent variable:
	Average course grade (GPA)
Workload	-0.361^{***}
	(0.101)
Helpfulness and easiness	0.377^{***}
	(0.109)
Easiness	0.028^{*}
	(0.016)
Easiness and perceived difficulty	-0.466^{*}
	(0.280)
Constant	3.057^{***}
	(0.122)
Observations	106
R^2	0.296
Adjusted R^2	0.268
Residual Std. Error	0.292 (df = 101)
F Statistic	10.625^{***} (df = 4; 101)
Note:	*p<0.1; **p<0.05; ***p<0.01



Results: The parameters are jointly significant per the overall significance F-test results:

$$F\text{-statistic} \approx 10.625 > F\text{-critical}_{4,101} \approx 2.46 \Rightarrow \textit{Reject}$$

Interpretation: An increase in workload rating by 1 point decreases average course grades by approximately 0.361 of a grade point. When a course is rated as easy and the instructor is rated as helpful, average course grades show an apprxoimate 0.377 increase in grade points for a corresponding 1-point increase in helpfulness and easiness. Courses rated to be easier indicate an approximate 0.028 increase in average course grades. The last parameter, for the interaction between the easiness rating and percentage of P/NP scores taken, has a value of -0.466 and implies a decrease in average course grades by 0.466 of a grade point. As the percentage of P/NP scores taken is interpreted to reflect the perceived difficulty of the course for this data set, the interaction may be understood as students' anticipation of course difficulty versus actual difficulty (as measured near the end of the course).

Analysis: More difficult courses (as reflected by higher workloads) yield lower average course grades for non-STEM courses. Previously, Model 3 indicated that STEM courses with higher workload apparently correlated with higher average course grades. This distinction is interesting to note, and raises the question as to whether more difficult courses contain higher workloads, and whether this incresaed difficulty (if applicable) motivates students to study harder and thus achieve higher course grades. It is also probable, however, that more difficult courses are also prone to grade inflation, so that despite the higher workloads, average course grades remain high. While Model 3 may support this reasoning, this model seems to suggest rather that courses with greater workloads are less likely to be grade-inflated.

The latter effect as demonstrated by this model would be logical in the context of lower-division courses, oftentimes of which tend to be prerequisites for admission into a particular major or minor. These courses generally have higher workloads and give lower grades and worse curves, so that a certain portion of students are "weeded out" of the respective major or minor.

The parameter for the percentage of scores taken as P/NP also highlights an interesting phenomenon: it appears that perceived difficulty, correlating negatively with average course grades, on behalf of students are quite accurate. In other words, it seems that students have fairly accurate sentiments regarding the difficulty of a course, and that those who opt to take a P/NP grade would likely receive a low grade in the course if taken for a letter grade.

Model 5: Percentage of As given on overall course rating It is reasonable to predict that a significant proportion of students would be concerned with their likelihood of receiving an A grade in a particular course. This regression model tests the direct effect of the chances of receiving an A in a given course and the overall course rating:

$$\ln(OVERALL) = \zeta_1 + \zeta_2 PercA + e, \tag{12}$$

where $PercA$ gives the percentage of students who receive an A grade in a course and $OVERALL$ reflects the average overall course rating on a scale of 1-5.

The following hypothesis was constructed to test for the significance of $PercA$:

$$H_0 : \zeta_2 = 0 \qquad\qquad H_1 : \zeta_2 \neq 0$$

Results are summarised in the table below.

Table 8: Percentage As given and overall rating

	Dependent variable:
	Logged overall rating
Percentage As given	0.275^{**}
	(0.106)
Constant	1.197^{***}
	(0.045)
Observations	106
R^2	0.061
Adjusted R^2	0.052
Residual Std. Error	0.223 (df = 104)
F Statistic	6.713^{**} (df = 1; 104)
Note:	*p<0.1; **p<0.05; ***p<0.01

Results: The parameter is significant per the t-test results:

$$t\text{-statistic} \approx 2.591 > t\text{-critical}_{0.975,104} \approx 1.983 \Rightarrow Reject$$

Interpretation: Courses with a greater percentage of students receiving A grades show an increase in overall course rating by approximately 27.5%, statistically significant at the 1% level.

Analysis: As expected, courses with a more generous curve (i.e., where a higher proportion of students receive A grades) are rated better than courses with less generous curves. Based on this relationship, one can reasonably deduce that there indeed is an incentive to grade-inflate (*see subsection*: **The inflation incentive**). An instructor who gives a greater percentage of A grades for a particular course very likely will receive substantially higher overall ratings for their course. This model strongly supports one of the reasonings put forth in explanation of grade inflation incentives.

Model 6: Course overall and feature ratings on percentage of As given As the percentage of A grades significantly affects the overall course rating, the converse effect may also be modelled and tested for significance:

$$PercA = \eta_1 + \eta_2 EASN + \eta_3 WKLD + \delta_1(STEM \times percPNP) + \delta_2(OVERALL \times STEM) + e, \quad (13)$$

where $PercA$ gives the percentage of students who receive an A grade in a course, $EASN$ gives how lenient the course and/or professor is perceived to be, $WKLD$ reflects how much work the course consisted of, $percPNP$ denotes the perceived difficulty of the course by students, $STEM$ indicates the presence of a STEM course, and $OVERALL$ reflects the average overall course rating on a scale of 1-5.

The following hypothesis was tested for overall significance:

$$H_0 : \eta_k = 0 \text{ where } k = 1, 2, 3;\ \ \delta_k = 0 \text{ where } k = 1, 2 \qquad H_1 :\ \geq \text{ one } \eta_k \neq \text{ and/or } \geq \text{ one } \delta_k \neq 0$$

Results are shown in the statistical summary table below.

Table 9: Course feature ratings and percentage As given

	Dependent variable:
	Percentage As given
Easiness	0.247^{***}
	(0.053)
Workload	-0.143^{**}
	(0.059)
STEM perceived difficulty	-1.836^{***}
	(0.642)
STEM overall rating	0.030^{***}
	(0.011)
Constant	0.103
	(0.066)
Observations	106
R^2	0.312
Adjusted R^2	0.284
Residual Std. Error	0.173 (df = 101)
F Statistic	11.425^{***} (df = 4; 101)
Note:	*p<0.1; **p<0.05; ***p<0.01

Results: The parameters are jointly significant per the overall significance F-test results:

$$F\text{-statistic} \approx 11.425 > F\text{-critical}_{4,101} \approx 2.46 \Rightarrow \textit{Reject}$$

Interpretation: An increase by 1 point in the easiness rating increases the percentage of A grades given by approximately 0.247%. A 1-point increase in the workload rating decreases the percentage of A grades given by approximately 0.143%. STEM courses perceived to be more difficult by students yield a decrease of approximately 1.836% in the percentage of As given, while a higher overall rating for STEM courses corresponds to an approximate 0.030% increase in the percentage of As given.

Analysis: This regression model is based upon the assumption that students select courses based on past course evaluations, which is reasonably likely. Likewise, course instructors are able to view past course evaluations, and adjust grading schemes accordingly.

Model 7: Percentage of As given by tenured vs. untenured instructors and total pay A common explanation for inflating grades arises from the fact that promotion to tenure status depends substantially upon student course evaluations. According to this theory, untenured instructors are more likely to inflate grades in order to earn higher course evaluation scores from students. As such, the following log-linear multiple regression tests for potential correlations:

$$\ln(TOTALPAY) = \kappa_1 + \delta_1(LECT \times PercA) + \delta_2(PROF \times PercA) + e, \tag{14}$$

where $TOTALPAY$ reflects the annual total pay in USD, $PercA$ gives the percentage of students who receive an A grade in a course, $LECT$ indicates an untenured instructor, and $PROF$ indicates a tenured instructor.

The following hypothesis was tested for overall significance:

$H_0 : \kappa_1 = 0;\ \delta_k = 0 \text{ where } k = 1, 2$ $\qquad$ $H_1 : \kappa_1 \neq 0 \text{ and/or } \geq \text{ one } \delta_k \neq 0$

Results are summarised in the table below.

Call: lm(formula = log(total_pay) ~ lecturer:PercA + prof:PercA, data = ugrad)

Residuals: Min 1Q Median 3Q Max -1.12641 -0.31520 0.03375 0.27769 1.32308

Coefficients: Estimate Std. Error t value Pr(>|t|)
(Intercept) 11.78537 0.06864 171.705 < 2e-16 ***lecturer:PercA -0.73303 0.22491 -3.259 0.00157*** *PercA:prof 1.02533 0.21647 4.737 7.97e-06* ** — Signif. codes: 0 ‘**’ 0.001** ’’ *0.01* ” 0.05 ‘.’ 0.1 ’ ’ 1

Residual standard error: 0.4372 on 91 degrees of freedom (12 observations deleted due to missingness) Multiple R-squared: 0.3712, Adjusted R-squared: 0.3574 F-statistic: 26.86 on 2 and 91 DF, p-value: 6.8e-10

Table 10: Percentage As given and total pay

	Dependent variable:
	Logged annual pay (USD)
Untenured	-0.733^{***}
	(0.225)
Tenured	1.025^{***}
	(0.216)
Constant	11.785^{***}
	(0.069)
Observations	94
R^2	0.371
Adjusted R^2	0.357
Residual Std. Error	0.437 (df = 91)
F Statistic	26.859^{***} (df = 2; 91)
Note:	*p<0.1; **p<0.05; ***p<0.01

Results: The parameters are jointly significant per the overall significance F-test results:

$$F\text{-statistic} \approx 26.859 > F\text{-critical}_{2,91} \approx 3.12 \Rightarrow Reject$$

Interpretation: Untenured instructors who give more A grades have total annual pay of approximately 73.3% less, while tenured professors who give the same share of A grades have a total annual pay increase of approximately 102.5%.

Analysis: It appears that untenured instructors indeed earn lower salaries than tenured instructors. This reasoning may be incomplete, however, since untenured instructors (especially lecturers) may have other paid positions, such as teaching at another college. Thus, the total pay for lecturers may not be reflective of their actual total annual salary, and their hours may be uncomparable to those of full-time tenured professors. On average, nevertheless, untenured instructors do seem to make less than tenured professors.

Model 8: Course feature ratings on percentage of As given, STEM vs. non-STEM This final regression model also analyses the factors influencing the percentage of A grades given. A parameter for (tenured) professors rated as more lenient is added to this model to explore the likeliness of such professors to give more A grades, along with parameters for STEM course features. The multiple linear regression may be represented by

$$PercA = \theta_1 + \delta_1(PROF \times EASN) + \delta_2(STEM \times EASN) + \delta_3(STEM \times percPNP) + e, \tag{15}$$

where $PercA$ gives the percentage of students who receive an A grade in a course, $PROF$ indicates a tenured instructor, $percPNP$ denotes the perceived difficulty of the course by students, $EASN$ gives how lenient the course and/or professor is perceived to be on a scale of 1-5, and $STEM$ indicates a STEM course.

The following hypothesis test was conducted to determine overall significance:

$$H_0 : \theta_1 = 0;\ \delta_k = 0 \text{ where } k = 1, 2, 3 \qquad\qquad H_1 : \theta_1 \neq 0 \text{ and/or } \geq \text{ one } \delta_k \neq 0$$

Results are shown in the following statistical summary table.

Results: The parameters are jointly significant per the overall significance F-test results:

$$F\text{-statistic} \approx 5.108 > F\text{-critical}_{3,90} \approx 2.72 \Rightarrow Reject$$

Interpretation: Tenured professors rated as more lenient tend to give approximately 0.040% more A grades, while STEM courses rated as easier yield approximately 0.059% more A grades. Au contraire, STEM courses perceived to be more difficult are likely to have approximately 2.000% less A grades.

Analysis: Many of the previous patterns seem to be reaffirmed in this set of regression results. Intuitively, as implied by the regression parameter interpretations, easier courses tend to give out a greater share of A grades than more difficult courses. An interesting area to further explore may be whether courses are intentionally made easier in order to allow a greater percentage of students to "earn" A grades.

Implications

If monetary currency is allowed to fluctuate and continuously increase, one can argue that grades may well be permitted to follow suit - an A today can just become an AA tomorrow and AAA the next, until eventually everyone who starts with an A^n winds up with an A^{n+1} in the subsequent grading period. In fact, this is precisely what economist Tim Harford proposes: converting grade "distortion" to true grade inflation by uncapping the upper GPA limit[15]. Admissions officers, firms, and other gatekeepers can then look forward to

[15] Harford, Tim. *Outside Edge: An easy answer to grade inflation.* 21 March 2009.

Table 11: Course easiness and percentage As given

	Dependent variable:
	Percentage As given
Professor easiness	0.040**
	(0.015)
STEM course easiness	0.059***
	(0.021)
STEM expected difficulty	−2.000***
	(0.753)
Constant	0.299***
	(0.028)
Observations	94
R^2	0.145
Adjusted R^2	0.117
Residual Std. Error	0.195 (df = 90)
F Statistic	5.108*** (df = 3; 90)
Note:	*p<0.1; **p<0.05; ***p<0.01

obtaining a dual degree in economics in order to learn the techniques for properly deflating these outlandish grades, as an economist would periodically deflate currencies.

The essential issue at stake, however, does not revolve around exactly how many pluses trail after your A, but rather the intrinsic value of grades. Since grades are the currency in the educational market which function to reveal student ability - a vital hidden type - to potential employers, artificially assigning a high grade to every student in effect attenuates this signalling value of grades.

A classic signalling game[16] spells out this situation, comprising three players: students of ability types high (H) and low (L), universities assigning students grades of type high (h) and low (l), and firms hiring students for jobs of type good (g) and bad (b). Since student ability is a hidden type, firms must thus rely upon grades given by universities, which are presumed to reveal this hidden information. Universities essentially function as matchmakers: the goal of the game is to match H-type students with g-type jobs[17] in order to maximise market output.

A problem arises when our matchmakers are not so honest and throw the rules out the door in assigning every student an h grade and thus sending wrong signals to firms. Front and center: grade inflation.

Job misalignment - the direct result of this grade misassignment (*i.e.*, overallocation of h grades) - leads to market inefficiency and deadweight loss. The game winds up as one of negative-sum: overall payoffs are lower due to market inefficiency in equilibrium, albeit individual agents may benefit. The aforementioned incentives to grade-inflate by individual schools and instructors attest to this outcome. Each individual school has no incentive to not grade-inflate, as doing so would, in its flawed fantasy, allow L-type students to (also) land g-type jobs, and hence improve the reputation of the respective school. But in doing so, the influx of L-type students into the pool of H-type students in effect "waters down" the quality of the collective

[16] This concept of modelling grade inflation as a signalling game is largely inspired by a working paper elaborating on the derivatives and economic models behind such a fitting depiction. See Yang, H. & Yip, C.S. (2003). *An Economic Theory of Grade Inflation*. Department of Economics, University of Pennsylvania. *Working paper, first draft April 2002.*

[17] The complexities surrounding job assignment in the signalling game are further detailed in Chan, W., Hao, L., and Suen, W. (2007). [*A Signaling Theory of Grade Inflation*] (https://onlinelibrary.wiley.com/doi/full/10.1111/j.1468-2354.2007.00454.x). *International Economic Review*, *48*(3).

reputation of H-type students - the public good. As with any good, a decrease in quality logically leads to a decline in prices. In this case, the price - the market wage for H-type students in g-type jobs - is reduced (*i.e.*, previously high-paying g-type jobs no longer pay as well, and become b-type jobs). Hence we can say that bad jobs (of b-type) replace good jobs (of g-type) in long-run competitive equilibrium.

Conclusion

Altogether, the ten total regression models constructed with undergraduate-level course data from the University of California, Los Angeles indicate strong support for the case of significant grade inflation (or, more accurately interpreted as grade compression). Specifically, the Latin honours eligibility thresholds for both the College of Letters and Sciences and the Henry Samueli School of Engineering and Applied Sciences precisely illustrate the phenomenon of grade compression, where higher grades rise significantly slower than lower grades, until all approach the current upper GPA cap of 4.0. The eight subsequent regressions, built with undergraduate-level grade distribution and course evaluation data, provide strong evidence in support of prevalent grade inflation in postsecondary education, especially at institutions with higher selectivity. When taken together with the representation of grades as the currency of the educational market within a classic signalling game, the regression results may reflect the potential presence of grade signalling value attenuation through skewed signals, grade misassignment, and job misalignment.

To retrace our trajectory back onto the search for the grade multiplier, an interesting recapitulation may be to rephrase our original question regarding the value of this multiplier to one of its sign: could it be the case that our multiplier, in actuality, is negative?

The posited perspective of (probable) attenuation in grade signalling power would effectively prove the case of a negative grade multiplier: market inefficiency, arguably, may be considered contractionary.

The previously postulated long-run competitive equilibrium outcome likewise illustrates the effects of our grade multiplier. Here lies an unhappily-ever-after, where good intents translated into less-than-good output - the flood of unsubstantiated As merely drowned out the substantiated ones, which may be quite aptly denoted by Gresham's law of inflation, wherein the bad (grade) drives out the good.