An Analysis of Grading Patterns in Undergraduate University Courses

Gary M.Weiss, Luisa A. L. Rosa, Hyun Jeong and Daniel D. Leeds
Department of Computer & Information Sciences, Fordham University, USA
{gaweiss, lrosa6, sjeong18, dleeds}@fordham.edu

***Abstract*— University undergraduate course grades have several purposes: they provide feedback to the student and motivation to perform well; serve as admission criteria for entering a major; and are used as selection criteria for future employers and graduate programs. Accurate assignment of grades is therefore important and critical to ensure fairness. However, grades may also impact the student's assessment of the instructor, which leads to a conflict of interest when such assessments are a component of employment, salary, or tenure decisions. This paper performs a detailed descriptive analysis of undergraduate grades collected over an eight year period from a major metropolitan university. Interesting grading patterns are identified and discussed, and the analysis suggests that grading policies vary substantially at the department, course, and instructor level. A connection is observed between course/department enrollment and average grades assigned. A particular focus of this study involves describing the grading behavior of instructors, with the goal of identifying instructors that assign grades that are statistically far above or below the norm. The analysis performed in this study can be applied to grade data from other universities using our publicly available Python-based analytics tool. The results of these analyses can be used to better understand existing grading policies, identify potential sources of grading inequities, and, when appropriate, take corrective action.**

Keywords—education, grading, data analysis, fairness

I. Introduction

University course grades are important, of great concern to students, and serve several purposes [1]. First and foremost, they are informational and measure the student's performance in the course and knowledge of the material. They are also used for placement into courses, assigning or continuing scholarships, admission into majors and future degree programs, and as a component of employment decisions. Grades that accurately reflect a student's learning are therefore required for fairness. Grades also serve as a form of motivation, as research shows that stricter grading standards [6], competition [4], and letter versus pass/fail grades [8][10] are associated with improved learning. One study of an undergraduate accounting class [6] showed that when the translation from numerical to letter grade was made stricter, exam performance improved; students subject to the stricter scale would achieve 20% more A's/B's than those subject to the more lenient grading scale if final grades were both assigned using the more lenient scale. Consequently, lenient grading may negatively impact learning.

There are impediments to accurate grading. Grades may impact a student's assessment of the instructor, which leads to a conflict of interest when student assessments are used for instructor employment, salary, and tenure decisions. This can lead untenured faculty and adjuncts with limited job security to inflate student grades. The linkage between higher grades and higher instructor assessment is supported by a study that showed that if students are given grades one standard deviation above, versus below, the mean, then the instructor's student ranking increases by one full standard deviation [9]. Numerous studies also establish a relationship between faculty rank and grading leniency. One study showed that adjunct faculty assign the highest grades [15], while another study showed that untenured and part-time adjunct faculty assign higher grades than tenured faculty [12]. Although the latter study concluded that high grades are used to obtain better assessments, the authors postulated that higher grades could be due to untenured and adjunct faculty having less experience and ability to distinguish between differing levels of learning. This linkage between stricter grading and experience is supported by another study that showed that untenured faculty grade more strictly over time and raise their grading standards as they approach tenure [7].

This paper performs a descriptive data analysis of eight years of undergraduate grades at a large metropolitan university, with the goal of finding interesting grading patterns and grading patterns that *may* suggest that the assigned grades are not accurate (i.e., deviate from grades from other instructors and courses). The data and grading patterns are analyzed at the department, course, and instructor levels. Key findings are that grades are heavily dependent on the academic department offering the course and that some instructors assign grades far outside the norm. The grading analysis performed in this study is implemented in a publicly available, open-source, tool [14] that will permit other academic institutions to perform similar analyses. We believe the analyses described in this paper, and implemented in our tool, can lead to meaningful improvements in grading and hence a more fair and effective grading system.

II. Dataset Description

Our data set is based on eight years of undergraduate data from Fordham University, located in New York City with an enrollment of 9,000 undergraduate and 7,000 graduate students. Each record represents one student in a specific course section and includes: student and instructor ID, course name and number, course department, term, and student grade using a 0.0 (F) - 4.0 (A) scale. Table I provides key dataset statistics. Student identifiers were anonymized and course sections with fewer than five students were omitted to enhance privacy. Even with such measures, the data is too sensitive to be shared.

TABLE I. Summary Data Set Statistics

Feature	Unique Values
Record Number	442,230
Student ID	24,654
Instructor ID	2,195
Course Name & Number	2,505
Course Section	21,504

III. Analysis

This study analyzes grading trends at the student, instructor, course, and department levels.

A. General Grading Trends

We begin by analyzing the impact of student year (freshman to senior) and course level (1000 – 4000) on student grades. Table II shows how each of these factors relate to student grades independently and in combination. Certain entries are very common (30% of all enrollments are associated with freshman in 1000 level courses) while others are very rare (0.03% are associated with freshman taking 4000 level courses).

TABLE II. Average Grade by Student Year and Course Level

	Course Level				
Student Year	***1000***	***2000***	***3000***	***4000***	***Average***
Freshman	3.109	3.276	3.037	3.235	**3.122**
Sophomore	3.190	3.284	3.217	3.257	**3.224**
Junior	3.169	3.313	3.281	3.322	**3.264**
Senior	3.182	3.348	3.333	3.411	**3.331**
Average	**3.137**	**3.305**	**3.275**	**3.389**	

The last column of Table II indicates a monotonic increase in average grade over the four years a typical student spends at the university; this pattern is replicated for each course level from 2000 to 4000. This increase can be explained by increasing maturity levels, improved study habits, attrition of weaker students, and a reduced focus in later years on weeding out weak students. The slight decrease in 1000-level course grades from sophomore to later years may result from only weaker students taking such courses in their final years. Our results are consistent with prior research that showed that freshman and sophomore courses have lower grades than junior and senior courses [13]. The final row of Table II shows that grades also increase with increasing course level, although the pattern is not observed between 2000 and 3000 level courses. This pattern may be due to higher level courses being taken mainly by students later in their academic career; however, it may also be due to students taking more courses that interest them rather than required core courses. The drop in grades from the 2000 to 3000 level courses may be due to the prevalence of 3000 level core courses. All these patterns may impact other patterns that we observe.

B. Student-Level Grading Trends

Fig. 1 shows the distribution of student GPA's, weighted by the number of course credits, using the 7400 students that completed at least 70% of their required degree credits within the data set. Dashed vertical lines identify the 25th, 50th, and 75th percentiles. Sixty-five students had a weighted GPA below 2.0 and are not represented in the figure. The median GPA is 3.31, while 25% of the students have GPAs above 3.57 and 25% below 3.02. The fact that only 25% of the students have a GPA below 3.02 shows the impact of grade inflation, as the faculty handbook defines a 3.0 (B) as "Good; solid and *above average* level" and a 2.33 (C+) as *average* performance. The skewed distribution causes the mean student grade of 3.25 to fall below the median grade of 3.31.

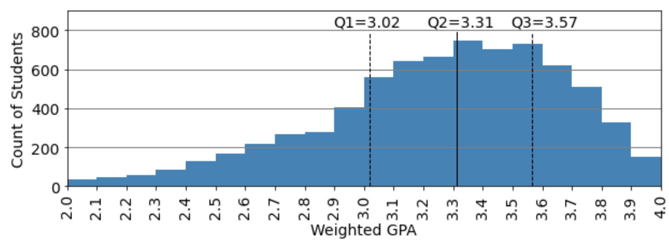


Fig. 1. Distribution of student weighted GPAs

C. Department-Level Analysis

University faculty are partitioned into departments that each support one or more majors. Fig. 2 shows the GPA for each department, based on all grades assigned by the department and weighted by course credits. Departments are placed into three categories: Arts, Humanities, and Languages; Communications and Social Sciences; and STEM (Science, Technology, Engineering, and Math). The departmental GPAs vary from a low of 2.83 (Chemistry) to a high of 3.73 (Urban Studies), with a mean of 3.32, denoted in Fig. 2 by a dashed horizontal line. STEM departments tend to have lower GPAs than departments belonging to the other two categories; all STEM departments have GPAs below the mean departmental GPA. These results agree with those from a prior study that found that grades are tied to course discipline and "courses emphasizing quantitative and factual learning tend to have assigned lower grades" [13]. A study of social science departments found that departmental grade differentials fall within a 7% range, and that "instructors that teach in more than one department grade more generously in departments that award higher grades, suggesting that grading differential policy is set by departments" [2].

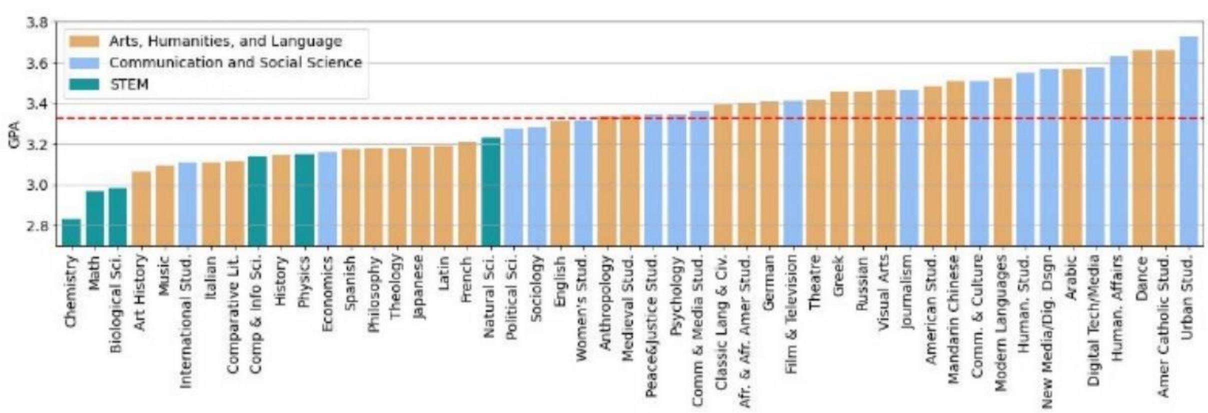


Fig. 2. Weighted GPA by department.

One of our hypotheses was that departments with lower enrollments assign higher grades because they are concerned about "scaring away" students or are reluctant to penalize students with whom the instructors have closer bonds due to the smaller department size. The scatter plot in Fig. 3 investigates the relationship between department size and department GPA; each point represents one department and is color coded using the same scheme as in Fig. 2. The cluster of points in the circle at the upper left corner clearly shows that there are many departments with high GPAs and low enrollments (i.e., below 5000); the colors indicate that none of these are STEM departments. There are a few departments with enrollments above 5000 that have GPAs above the department mean, but none of these have GPAs above 3.5—while 11 of the smaller departments have GPAs above this level. Of the 21 departments with the highest GPAs, 19 have enrollments below 5000 and reside in the circle. This establishes a strong relationship between departments with high GPAs and low enrollments. In contrast, there is no clear relationship between lower department GPAs and enrollments. The departments with the lowest GPAs are STEM departments, and they all have moderate enrollments. The data suggests that departments with very low enrollments may feel unusual pressure to assign higher grades or provide more effective instruction.

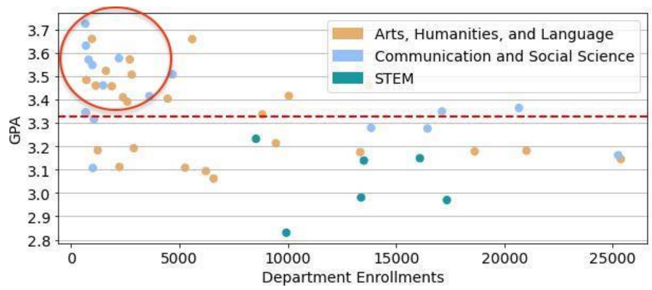


Fig. 3. Relationship between department enrollment and GPA.

D. Course-Level Analysis

We analyzed grading trends for individual courses. Fig. 4 shows the GPA of the 27 courses in the data set with at least 70 sections. Course grades tend to follow the grades associated with the offering department (e.g., STEM courses tend to assign low grades). The majority of these courses, including the lowest ten GPA courses, have lower average grades than their respective departments, which supports a connection between high enrollment courses and lower course grades.

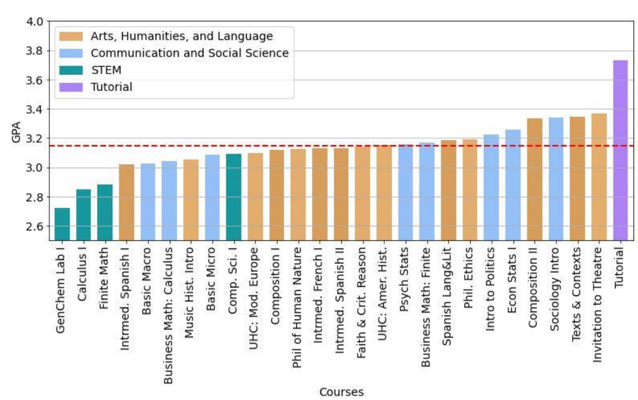


Fig. 4. GPAs for popular courses.

The tutorial course has the highest grades. This course is not associated with any department but is used when there are not enough students to run a regular class. The high grades may occur because it is difficult for instructors to assign low grades when they have more contact with a student or because it is difficult to characterize a grade distribution with few students and this makes it is difficult to assign low grades; alternatively it may be due to better student learning in small classes.

We next analyze the grade distributions associated with each of the 221 courses with a total enrollment of at least 300 students. A grade distribution vector is computed for each course, where each of the ten elements in the vector correspond to the percentage of the students receiving the corresponding letter grade. The k-means clustering algorithm was applied to these course grade vectors using k=4 clusters (the elbow and silhouette methods were used to set k=4). Each cluster is described in Table III. The C+, C, and C- grades are combined to conserve space and GPA represents the average grade.

TABLE III. AVERAGE COURSE GRADE DISTRIBUTION BY CLUSTER

	Letter Grade (%)									
Cluster	***A***	***A-***	***B+***	***B***	***B-***	***C***	***D***	***F***	***GPA***	***Count***
0	27.7	14.6	13.2	14.2	8.8	16.1	3.1	2.3	3.11	58
1	40.6	23.0	13.6	10.4	5.0	5.6	0.7	1.0	3.49	47
2	20.6	22.9	20.0	16.0	8.6	9.6	1.2	1.2	3.25	71
3	12.9	12.9	15.6	18.3	13.3	22.6	2.7	1.6	2.91	45

Below is a simplified description of the clusters:

- Cluster 0: "A" is the most common grade, with the next three grades all having lower, but similar, frequencies.
- Cluster 1: The highest GPA with many A's and a generally decreasing trend of letter grade frequencies.
- Cluster 2: The three top grades have similar frequencies.
- Cluster 3: Only cluster with a bell-shaped distribution, which peaks at "B".

While A's are the most common grade for clusters 0 and 1, cluster 0 has a substantially lower GPA because it contains many more low grades. Comparing cluster 0 and cluster 2, which have similar GPAs, shows an interesting difference: cluster 0 has substantially more A's and fewer A-'s than cluster 2 (cluster 0 also has nearly twice as many A's as A-'s while cluster 2 has slightly fewer A's than A-'s). Cluster 0 also has many more D's and F's. As mentioned earlier, cluster 3 is the only cluster that exhibits an approximately normal distribution and is associated with the lowest GPA. These findings indicate that a scalar value, like mean GPA, may be insufficient for understanding grading practices. Different grading distributions will also have a non-uniform impact on students based on their different levels of academic ability. Although student performance will vary based on subject, a student in the top 5% academically is likely to receive an "A" in courses belonging to any of the clusters, while the grades of those students near the very bottom may vary widely based on which cluster the course is associated with.

Table IV shows the distribution of high-enrollment courses by cluster for a variety of departments. Cluster values that cover a substantial fraction of the courses are in boldface. Departmental courses are often concentrated in one cluster. The traditional science departments (Biology, Chemistry, Physics) are the only ones with more than half of their courses in cluster 3, which is the only cluster with a bell-shaped grade distribution (it also has the lowest GPA). This observation reinforces our prior finding that STEM courses have lower grades. Mathematics is the only STEM subject with no courses in Cluster 3 as every popular class falls into cluster 0.

TABLE IV. COURSE CLUSTER DISTRIBUTION BY DEPARTMENT

	Cluster				
Department	***0***	***1***	***2***	***3***	***Total***
Biological Sciences	1	1	**5**	**7**	14
Chemistry	0	0	2	**8**	10
Comm & Media Studies	2	**8**	**12**	1	23
Comp. Info Science	**7**	1	2	3	13
Economics	**8**	1	2	3	14
English	0	0	**5**	0	5
History	1	1	**6**	2	10
Mathematics	**13**	0	0	0	13
Natural Science	4	6	2	4	16
Philosophy	0	0	**3**	0	3
Physics	1	2	0	**6**	9
Psychology	**6**	**8**	1	0	15
Spanish	1	0	**4**	1	6
Theology	0	0	**12**	2	14

The departmental differences in Table IV could be driven by the departmental GPA differences shown previously in Fig. 2. Fig. 5 uses parallel coordinates [11] to explore this further and determine the consistency of grading patterns for courses in Mathematics, Theology, and Chemistry (these departments were selected because they each have most of their courses concentrated in a single, different, cluster). Each line corresponds to a single course and represents the grade distribution ("D" and "F" grades are rare and omitted). Mathematics and theology courses follow a consistent pattern that matches the pattern for their department's cluster, but the theology courses seem to exhibit two different patterns for the A's/A-'s. Chemistry courses show a less consistent pattern. These grading patterns are interesting and warrant further study.

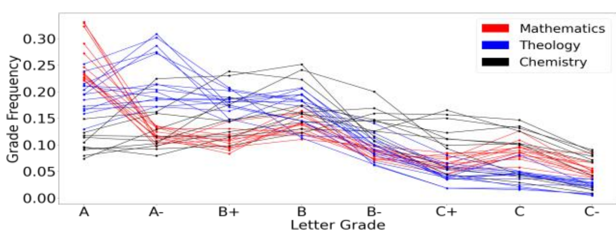


Fig. 5. Grade distribution for three deparments.

E. Instructor-Level Analysis

This section looks at the grading practices of individual instructors. These practices can vary widely as there are no constraints, such as a limit on the percentage of A's. Fig. 6 shows that the average grade assigned per instructor yields a distribution that is symmetric, approximately normal, and has a mean of 3.22.

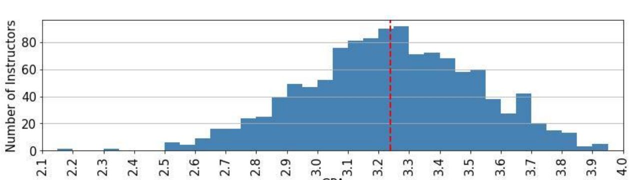


Fig. 6. Instructor GPA distribution (minimum six sections per instructor)

Grading inconsistency is a great concern as it raises issues of fairness and, as discussed earlier, can lead to students flooding the sections taught by "easy graders," reduced student effort, and diminished learning. If we can identify instructors that assign grades that deviate from the norm, they can be made aware of potential anomalous grading practices and take corrective action. We next examine the average course grades assigned by instructors for the same course. Large differences are unlikely to be due to differences in student ability but could be due to differences in instructor effectiveness.

1) Scatter plots of instructor grade distributions. Instructor grade distributions are provided in Fig. 7 for several courses with many sections and instructors. Each sub-figure displays the distribution of average grades per instructor for a specific course, aggregated over all course sections (each data point corresponds to one instructor). Standard deviation information is provided to identify instructors that assign extreme grades. The total course enrollment per instructor is provided on the y-axis so that we can focus more attention on the instructors with more reliable information and that have the most impact. Fig. 7 shows that the most extreme instructor GPAs often co-occur with low enrollments while high-enrollment instructors tend to have GPAs close to the mean, indicating that some extreme values may be due to small sample size. However, as we will see shortly, there are instructors with high enrollments that assign grades relatively far from the mean. The Fig. 7 caption provides more details about the figures.

2) Definition and description of distributional statistics. Excess kurtosis and skew is provided in Fig. 7 for each course to characterize these distributions and facilitate comparisons. The mean values are indicated by the value at standard deviation = 0. The kurtosis, skew, and mean values are based on the univariate average instructor grades and do not involve the enrollment values; hence when trying to understand the distribution one should only focus on the density of the points along the x-axis. We utilize the current format rather than a histogram because of our interest in the enrollment values.

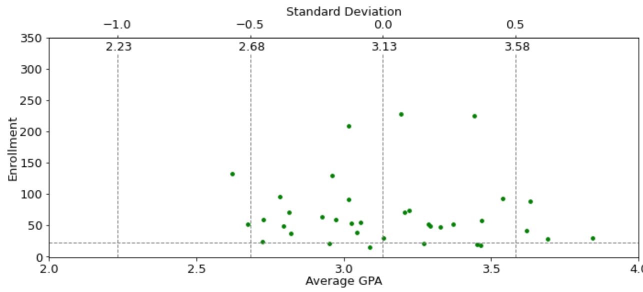
Skew [5] is a measure of the asymmetry of a distribution about its mean. A negative (positive) skew indicates that the tail is on the left (right) side and a skew of zero indicates a symmetric distribution. Kurtosis [3] measures the "tailedness" of the distribution and measures how close the distribution is to a normal distribution. We use Fisher's kurtosis and display the excess kurtosis (γ), which is defined as (Fisher's) kurtosis minus 3.0. Excess kurtosis can be interpreted as follows:

$\gamma < 0$: Platykurtic, thin tails (e.g., Uniform, $\gamma = -1.2$)

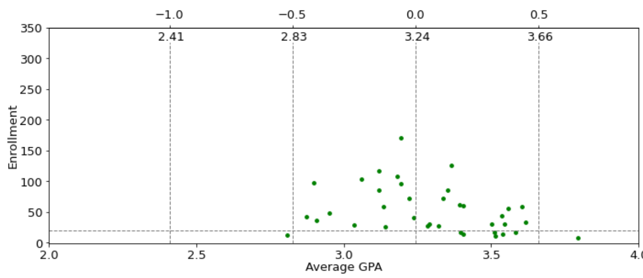
$\gamma = 0$: Mesokurtic: a normal (Gaussian) distribution

$\gamma > 0$: Leptokurtic, thick tails (e.g. Laplace, $\gamma = +3.0$)

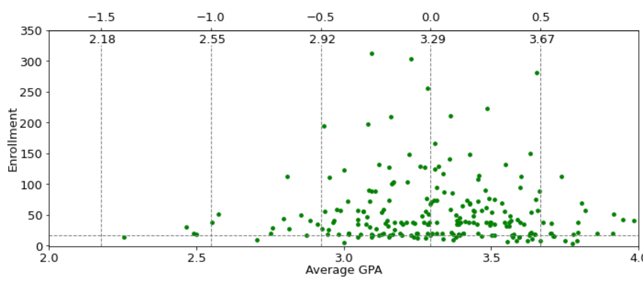
Five of the six distributions in Fig. 7 have negative excess kurtosis, indicating the average grades are clustered around the mean. The skew values are evenly split between positive and negative values. The highest skew (+0.56) is for "Finite Math," while the lowest skew (-0.46) is for "Texts and Contexts." This division of skews mirrors the department-level division in grading practices, with STEM courses assigning lower grades.



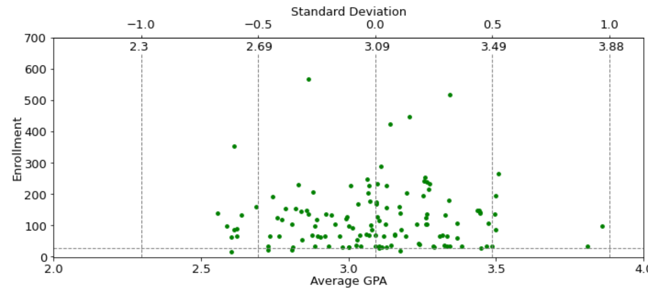
(a) Structures of Computer Science
Excess Kurtosis: -0.79, Skew: +0.25



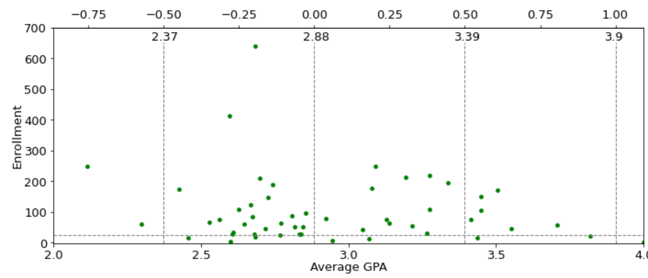
(b) Foundations of Psychology
Excess Kurtosis: -0.70, Skew: -0.21



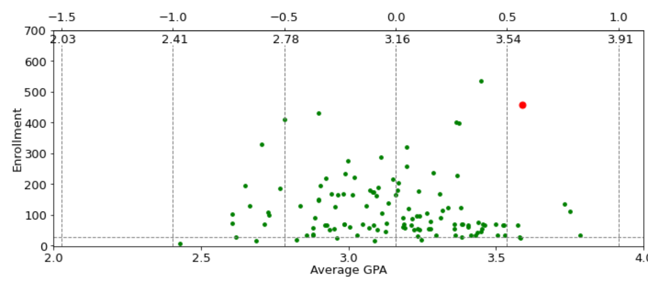
(c) Texts and Contexts (English)
Excess Kurtosis: +0.56, Skew: -0.46



(d) Philosophy of Human Nature
Excess Kurtosis: -0.06, Skew: +0.11



(e) Finite Mathematics
Excess Kurtosis: -0.15, Skew: +0.54



(f) Faith and Critical Reasoning (Theology)
Excess Kurtosis: -0.34, Skew: -0.14

Fig. 7. Instructor GPA scatter plots for specific courses versus total instructor course enrollment. Each point corresponds to one instructor and represents the average grade assigned in the course over all course sections the instructor taught (enrollment is the number of students taught across those sections). The top x-axis shows the standard deviation; the dashed vertical lines designate the ±0.5 and ±1.0 standard deviation boundaries. The y-axis ranges differ for the plots on the left and right sides. Each sub-figure caption provides the course name. The excess kurtosis and skew values for the univariate GPA distribution is provided. The Fisher definition of kurtosis was used; kurtosis measures the "tailedness" of the distribution, with a normal distribution having a zero value and positive (negative) values indicating heavier (lighter) tails than the normal distribution, showing a leptokurtic (platykurtic) distribution. Skew measures symmetry, with a negative (positive) skew having the distribution skewed to the larger (smaller) values, with a normal distribution being symmetric and having a skew of 0.

3) Identficiation of anomalous graders. Fig. 7 identifies the outlier graders and allows us to focus attention on those that teach many students and have the most impact. For example, the enlarged red data point in Fig. 7f shows that an instructor that taught 458 "Faith and Critical Reasoning" students assigned an average grade of 3.59 that is +0.57 standard deviations above the mean of 3.16. Most instructors would consider this disparity to be quite substantial. We do not see any high-enrollment instructors more than ±1.0 standard deviation from the mean, although quite a few with modest enrollments are ±0.5 away. The "Finite Math" course has an instructor with about 250 students and a mean grade under 2.2 that is about -0.75 standard deviations away from the mean. This is quite notable as a mean assigned grade of 2.2 is very low. These examples show that being even 0.5 to 1.0 standard deviations from the mean represents a meaningful outlier for grading, even though outliers traditionally are defined as more than three standard deviations from the mean.

4) Detailed analysis of instructors with anomalous grading. In this section we do a deeper analysis of instructors that assign grades more than a half standard deviation away from the mean. To best assess how much an instructor's grades deviate from those of other instructors, we exclude each instructor's grades from the mean used to calculate the z-score; we refer to this as the self-excluded z-score (Z_{se}). This analysis focuses on courses that have more than 25 different instructors and 1000 total enrollments. Only instructors with $|Z_{se}| > 0.75$ are considered; 76 unique instructors, covering 86 of 1562 (5.5%) instances, satisfy this condition. Table V shows information about the nine instructors that teach at least 100 students in the course in which their grades are out of the norm. Five of the instructors tend to give low grades and four tend to give high grades. The 86 instances were categorized using the course-level clustering model created in Section III.D. More than 80% of the negative-extreme instructors map to cluster 3, while all positive-extreme instructors map to cluster 1. This is not surprising since cluster 3 has the lowest GPA and cluster 1 has the highest GPA.

TABLE V. INSTRUCTORS WITH ANOMALOUS GRADING

Instr. ID	*Course*	Z_{se}	*Enrollment*	*GPA*
F78560	Composition II	-1.09	100	2.56
F11130	Composition II	-0.81	177	2.74
F27772	Econ. Statistics I	-0.80	117	2.60
F17127	Finite Math	-0.80	248	2.08
F30202	Intro to Sociology	-0.76	179	2.71
F72238	Faith & Crit. Reasoning	0.76	136	3.72
F13486	Econ Statistics I	0.78	124	3.90
F62351	Faith & Crit. Reasoning	0.79	111	3.74
F33259	Philosophical Ethics	0.85	195	3.85

Some instructors have interesting letter grade distributions. Instructor F27772's grades in "Econ Statistics I" are quite low and spread out, with 33% of students receiving A or A- and 20% D or F. This distribution, which yields a low GPA yet still assigns many high grades, is unique amongst the instructors analyzed in this section. A course this instructor taught that is not listed in Table V shows a similar distribution, so this distinctive grading pattern may be common for this instructor. Instructor F17127 taught 10 sections and 248 students of "Finite Mathematics" with an average grade of 2.08; the z-score is not the most extreme since the math department often assigns low grades. Grouping letter grades together for this instructor yields the distribution: A (9%), B (32%), C (30%), D (25%), F (5%). The instructor's section GPAs range from 1.84 to 2.36, with no clear temporal trend. The percentage of D's is quite concerning. This instructor also taught a similar course for business students and assigned 17% D's. The grade distributions for instructors with $Z_{se} > 0.75$ is less interesting since the average grades are so high that A and A- grades must dominate.

IV. CONCLUSION AND FUTURE WORK

This study provides an analysis of undergraduate grading data. The analysis found clear variations in grading between different academic departments, courses, and instructors. In particular, it was shown that departments with low total course enrollments tend to assign higher grades, STEM disciplines tend to assign lower grades, and there are substantial numbers of instructors that assign grades more than ±0.5 standard deviation from the overall course mean. These variations in grading practices raise issues of fairness and as described earlier, can lead to undesirable consequences. Grade distributions were also used to characterize individual instructors, courses, and departments. Notably, we found that instructors and departments often have distinctive letter grade distributions. We also noted that undergraduate courses often do not exhibit a traditional bell-shaped curve.

This study can help improve grading practices by providing university instructors and administrators with an understanding of common grading patterns and existing grading differentials that can lead to inequitable treatment of students. The analyses presented in this study can identify potential problems, such as an instructor that consistently gives unusually high or low grades, so that corrective action can be taken. As a second example, if a department identifies a course in which grades are well above or below the mean of other courses, they can reconsider how content is tested and grades assigned. The analyses can also be used to foster discussion at the department and institutional levels, which could lead to improved and more uniform grading policies. To bring these benefits to a wider community, our analyses have been incorporated into a software tool that is publicly available [14]. Our hope is that it will be used by educational researchers and practitioners.

While our work observes differences in grade distributions, we are not able to differentiate between high (low) grades from lenient (harsh) graders and those that result from more (less) effective instructors. It will be valuable to investigate these factors by considering learning effectiveness using student performance in future related courses [16] or via student surveys about instructor effectiveness. We also plan to compare the observed grading practices with those of other institutions, and have already collected much of the data needed for such an analysis. Our analyses can also be extended to consider instructor years of experience and rank, as there is a common belief, supported by research, that such factors impact grading [7, 12, 13, 16]. We would also like to combine our data with existing student surveys to extend existing research [9] on the impact of grades on student-based instructor assessment.

REFERENCES

[1] G. Adams and T. Torgerson, Measurement & Evaluation in Education, Psychology, and Guidance. New York: Holt, Rinehart & Winston, 1964.

[2] M. Beenstock, and D. Feldman, "Decomposing university grades: a longitudinal study of students and their instructors," Studies in Higher Education, vol. 43, no. 1, pp. 114-133, 2018.

[3] B.S. Chissom, "Interpretation of the kurtosis statistic," The American Statistician, vol. 24, No. 4, pp. 19-22, 1970.

[4] D.C. Clark, "Competition for grades and graduate student performance," Journal of Educational Research, vol. 62, pp. 351–354, 1969.

[5] D.P. Doane, and L.E. Seward, "Measuring skewness: A forgotten statistic?," Journal of Statistics Education, vol. 19, no. 2, 2011.

[6] F. Elikai, and P.W. Schuhmann, "An examination of the impact of grading policies on students' achievement," Issues in Accounting Education, vol. 25 no. 4, pp.677-93, 2010.

[7] S.E. Gohmann, and M.J. McCrickard, "Tenure status and grade inflation: A time series approach," Journal of the Academy of Business Education, vol. 2, pp. 1-8, 2001.

[8] R.M. Gold, A. Reilly, R. Silberman, and R. Lehr, "Academic achievement declines under pass-fail grading," Journal of Experimental Education, vol. 39, no 3, pp. 17–21, 1971.

[9] A.G. Greenwald, and G.M. Gillmore, "Grading leniency is a removable contaminant of student ratings," American Psychologist, vol. 52, no. 11, pp. 1209–1217, 1997.

[10] I.W Hales. P.T. Bain, and L.P. Rand, "An investigation of some aspects of the pass-fail grading system," Annual Meeting of the American Educational Research Association, New York, 1971.

[11] A. Inselberg, and B. Dimsdale, "Parallel coordinates: a tool for visualizing multi-dimensional geometry," Proceedings of the first IEEE conference on visualization: visualization '90, pp. 361-378, IEEE, 1990.

[12] M. Moore, and R. Trahan, "Tenure status and grading practices," Sociological Perspectives, vol. 41, no. 4, pp.775-781, 1998.

[13] J.E. Prather, and G. Smith, "A study of the relationships between faculty characteristics, subject field, and course grading patterns." Research in Higher Education, vol. 5, pp. 351–363, 1976.

[14] L.A.L. Rosa, H. Jeong, G.M. Weiss, D.D. Leeds. *Grade Analysis Tool* [Software], 2023. https://www.cis.fordham.edu/edmlab/software/grade-analysis-tool

[15] B.S. Sonner, "A is for 'adjunct': Examining grade inflation in higher education," Journal of Education for Business, vol. 76, pp. 5-8, 2000.

[16] G.M. Weiss, E. Brown, M. Riad-Zaky, R. Iannone, and D.D. Leeds, "Assessing instructor effectiveness based on future student performance," Proc. 15th Int. Conference on Educational Data Mining, Int. Educational Data Mining Society, Durham, UK, July 24-27, pp. 616-620, 2022.