Innovative Higher Education
https://doi.org/10.1007/s10755-019-09474-w

The Distribution of College Grades across Fields in the Contemporary University

Joseph C. Hermanowicz[1] **· David W. Woodring**[1]

Published online: 28 June 2019

Abstract

Scholars have argued that grade inflation is pervasive throughout colleges and universities and that it is presently at an all-time high. Inflation is, however, a temporal concept: it is theoretically impossible for grades to keep increasing on a fixed scale. In this article we examine a related, though empirically distinct, phenomenon: the *distribution* of grades across fields in a university. We question global statements about grade inflation and examine if and how the university grading structure is internally differentiated. We use the idea of *consensus*, the extent to which practitioners of a field agree, as a means to differentiate areas in a university. Based on undergraduate grade data from a large, public university in the U.S., we use cluster analysis to ascertain an "architecture" of grades. The results demonstrate significant variation in how grades are distributed across fields. The work identifies a need to probe further the linkages between field consensus, rigor, student learning, and grade allocation in college.

Keywords College students · Grades · Grade inflation · Academic fields · Cluster analysis

Background

Grades are a ubiquitous part of college. The dominance of grades in undergraduate life was established long ago in Becker, Geer, and Hughes' book (Becker, Geer, & Hughes, 1968) entitled *Making the grade.* College grades influence a student's graduation prospects, academic motivation, job choice, professional and graduate school selection, and access to scholarships (Rojstaczer & Healy, 2012, p.1). They also establish self-definition (Becker, Geer, & Hughes,

Joseph C. Hermanowicz is Professor of Sociology at the University of Georgia. He earned a Ph.D. from the University of Chicago. His research focuses on the sociology of higher education.

David W. Woodring is the Director of Social Problems and Solutions Group. He earned a Ph.D. from the University of Georgia. His interests center on program evaluation, research design and methodology.

✉ Joseph C. Hermanowicz
jch1@uga.edu

[1] Department of Sociology, The University of Georgia, 30602-1611 Athens, GA, United States

1968.) Institutionally grades are intended as markers of students' educational achievement (Pascarella & Terenzini, 2005). Colleges and universities customarily denote the qualitative equivalence of grades in student handbooks (A equals "excellent," B equals "good," C equals "average," D equals "poor," F equals "failing").

Assignment of grades is a principal responsibility of faculty members. College faculty members possess considerable freedom in establishing the criteria for and assigning grades. Indeed, academics in the United States consider autonomy in the assignment of grades to be a fundamental part of academic freedom (American Association of University Professors, 2015, pp. 29–30). As an occupational group, college teaching is generally a self-regulated profession, that is, its members socially control one another by determining rules and sanctions, and individual members command considerable autonomy. By the same token, self-regulated professions and organizations commonly have problems maintaining standards and ensuring ethical behavior (DeMarzo, Fishman, & Hagerty, 2005; Frey, 2006; Rojstaczer & Healy, 2012;). The assignment of grades does not occur in a vacuum. Rojstaczer and Healy argued, "Even if an instructor feels that grades have value and purpose, there are significant perceived incentives for that instructor to abandon any objective standard and award grades that are artificially high" (2012, p. 2). Krautmann and Sander (1999) presented evidence of the behavior wherein instructors "buy" better evaluations through more lenient grading.

The research literature on college grades strongly suggests that modern collegiate education transpires amidst widespread grade inflation (Hunt, 2008; Johnson, 2003; Milton, Pollio, & Eison, 1986). Grade inflation has increased over time, and it is presently at an all-time high (Rojstaczer & Healy, 2012). Forty-three percent of grades awarded across U.S. colleges and universities are A's (Rojstaczer & Healy, 2012). At public institutions 73% of grades are A's and B's, at private institutions, 86% (Rojstaczer & Healy, 2012). D's and F's total less than 10% of all letter grades (Rojstaczer & Healy, 2012). The average GPA at public schools is 3.0, at private ones, 3.3 (Rojstaczer & Healy, 2010). In 1960, as in the 1940s and 1950s, C was the most common grade nationwide; and D's and F's accounted for more grades combined than did A's (Rojstaczer & Healy, 2012, pp. 5–6).

One might conclude from the research that grading is easy across the board. Rojstaczer stated, "The bottom line is that grading nearly everywhere is easy. After 50 plus years of grade inflation across the country, A is the most popular grade in most departments in most every college and university" (Rojstaczer, 2016, p. 6). Grade inflation is measured by an increase in grades over time without an increase in student performance (Goldman, 1985; Hu, 2005; Rosovsky & Hartley, 2002). Inflation is temporal; there is a sustained increase in grades from one point in time to another.

Grading scales, however, have upper limits; and grades cannot be inflated infinitely. How far toward the high end of a scale will grades increase before a limit is reached? We seek to contribute to an understanding of college grades by investigating, not inflation, but a related, empirically distinct, phenomenon to account for grading practices: the *distribution* of undergraduate grades in a university. In this article we question global statements and summative conclusions about grade inflation. Sociologically, we turn attention to the possibilities of variation, that is, we examine if and how the university grading structure is *internally differentiated*. We are seeking to understand grading patterns better by looking deeper into institutions, rather than remaining at the institutional level as characteristic of previous research.

Further, a consideration of grade distributions within universities is relevant to ongoing concerns about rigor and actual student learning in college (Arum & Roksa, 2011). Variation in grade distributions within universities may be partial evidence in systematically different expectations and conventions (among fields) in assigning grades (e.g., it is easier to make an "A" in one field than another). To the extent it may exist, internal differentiation would help illuminate debates in which universal claims are made about grade inflation, declining rigor, and limited learning in college.

Codification and Consensus

As a way to examine internal differentiation of the university grading structure, we used the idea of field *codification*. A line of work on codification in science and education provides an avenue by which to examine the possible dynamics of grades by field (Becher, 1989; Braxton & Hargens, 1996; Hagstrom, 1965; Hargens, 1975; Hargens & Kelly-Wilson, 1994). Codification refers to "the consolidation of empirical knowledge into succinct and interdependent theoretical formulations" Merton & Zuckerman [1972] 1973, p. 507). Highly codified fields tend to be cumulative in their production of knowledge and, conversely, less codified fields more disparate in knowledge production. To this end, "research frontiers" are more discernible in highly codified fields whereas in comparatively less codified fields even the idea of research frontiers may be non-existent (Cole, 1993). Fields in the natural sciences and mathematics are generally associated with relatively high codification (e.g., Braxton & Hargens, 1996). Fields in the humanities and select social sciences are generally associated with relatively low codification (e.g., Braxton & Hargens, 1996).

An aspect of codification consists of *consensus*, the extent to which practitioners of a field agree upon, for example, problem choice, important research, suitable theory, valid research methods, definitions of career success—and arguably on assessment of students in courses in that field. Accordingly, *one would expect members of relatively high consensus fields to possess fairly clear definitions of success in their fields and correspondingly restrictive notions about the allocation of grades*. *One would expect members of relatively low consensus fields to possess comparatively more varied definitions of success in their fields and correspondingly relaxed notions about the allocation of grades.* Physicists who claim, for example, that members of the field (and correspondingly students in a course) can be rank ordered by achievement are testimony to high consensus (Hermanowicz, 1998, 2009). Members of other fields who claim that it would be difficult or impossible to rank order practitioners (and correspondingly students in a course) by achievement is evidence of low consensus (Hermanowicz, 1998, 2009). Again, generally, high consensus fields are those found in the natural sciences and mathematics; low consensus fields are found in the humanities and select social sciences (Braxton & Hargens, 1996).

By these dynamics, members of high consensus fields arguably offer the clearest and strictest judgments of work, whether their own or those of students. By contrast, members of low consensus fields arguably offer more plural judgments of their own and students' work. Consequently, we postulate that it is easier for students to achieve high grades in relatively low consensus fields (humanities and select social sciences) since success is defined in multiple ways. Correspondingly, we postulate that it is more difficult for a student to achieve high grades in relatively high consensus fields (natural sciences and mathematics) since success is defined in more restricted ways. Success is defined collectively according to a more absolute scale.

The Study

Purpose and Propositions

The purpose of our study was to explore patterns of grade distribution among subgroups within a large public university in the U.S. Grade distribution, hereafter called "mean GPA," was operationalized as the mean grade assigned to undergraduate students across courses within the university. Inductively, we sought grading patterns that are conceptually linked to field consensus. We analyzed grades assigned to undergraduates over a two-year period, and we found patterns within university subgroups of disproportionately high and low grades that generally coincided with areas of relatively high versus low field consensus.

Guided by the theoretic framework above, we articulated the following set of propositions.

- Proposition 1a: *Mean GPA across the university will mask significant internal differentiation in grade distribution*. We argue that such a summative statement as overall GPA conceals an internal architecture in the distribution of university grades.
- Proposition 1b: *Mean GPA will vary across departments*. We argue that fields do not grade with equivalent metrics and that this variation stems to a large degree from differences in field consensus.
- Proposition 1c: *Mean GPA will vary across colleges within a university.* We suggest that grades will be higher in pre-professional colleges, as opposed to an arts and sciences college, as a result of social- and self-selection effects (students meet additional criteria for enrollment in pre-professional colleges).
- Proposition 2a: *Within the arts and sciences (encompassing the core departments in a university), mean GPA will be higher in the social sciences and humanities than in other departments.* Low consensus fields are concentrated in the social sciences and humanities.
- Proposition 2b: *Within the arts and sciences, mean GPA will be lower on average in the physical and biological sciences and mathematics than in other programs*. High consensus fields are concentrated in the natural sciences and mathematics.

We proceed as follows. First, we describe the data and explain our analytic strategy. Second, we present the results and a discussion of the descriptive statistics, cluster analyses, and analysis of the distributional characteristics of grades associated with specific colleges and departments. Finally, we conclude by discussing the possible links between a grading structure of a university and field consensus.

Data Collection

Data come from grades at a major state university in the U.S. that enrolls approximately 30,000 undergraduates. The data were obtained in cooperation with the university's Office of Institutional Research. The university is the state "flagship" institution, and it has structures and students similar to other major public universities across the U.S. We use a major public university as a case that is illustrative of patterns in the distribution of grades across fields of study.

Grades for all undergraduates were obtained across two academic years: 2014–15 and 2015–16. This yielded 460,562 grades as points of interest. The university is organized into fifteen colleges and/or schools. The college of arts and sciences is the "core" college in the university, and it bears on several of the propositions. The college is organized into 28 departments and two schools, a school of art and a school of music. For present purposes we identify the schools as departments, resulting in 30 units. The list of colleges and units in arts and sciences are listed in Table 1. (For simplicity, our use of the term "field" corresponds to department and college.)

Data Analysis

We first used a *t*-test to investigate differences in mean GPA between two subgroups of departments within the college of arts and sciences (propositions 2a and 2b). For comparison we aggregated the individually assigned grades of undergraduates across all courses in departments in the social sciences, arts, and humanities ($n = 150{,}979$) as well as in departments in the physical and biological sciences and mathematics ($n = 106{,}436$) over the two-year period. While the ordinal grade data would normally require a nonparametric test (i.e., the grades are non-normally distributed), the large sample size rendered nonparametric tests inappropriate. In addition, given the large number of observations, we used a stratified random sample of grades from each group. To avoid assumptions about the underlying distribution of grades, we employed *t*-test with bootstrapping (1000 re-samples).

Table 1 University Schools/Colleges and departments

Colleges	Departments
1. Agriculture	1. Anthropology
2. Arts and Sciences	2. Art
3. Business	3. Biochemistry and Molecular Biology
4. Ecology	4. Biological Sciences
5. Education	5. Cellular Biology
6. Engineering	6. Chemistry
7. Environment and Design	7. Classics
8. Family and Consumer Sciences	8. Communication Studies
9. Forestry	9. Comparative Literature
10. Journalism	10. Computer Science
11. Pharmacy	11. Dance
12. Public Health	12. English
13. Public and International Affairs	13. Genetics
14. Social Work	14. Geography
15. Veterinary Medicine	15. Geology
	16. Germanic and Slavic Studies
	17. History
	18. Marine Sciences
	19. Mathematics
	20. Microbiology
	21. Music
	22. Philosophy
	23. Physics and Astronomy
	24. Poultry Sciences
	25. Psychology
	26. Religion
	27. Romance Languages
	28. Sociology
	29. Statistics
	30. Theater and Film Studies

Second, we used cluster analysis to explore the similarities and dissimilarities (i.e., internal differentiation) in grade distribution among colleges and departments based on their mean GPA. Specifically, to address propositions 1a-1c, we computed the mean GPA assigned to undergraduates within a set of colleges and departments to create two samples for analysis. We conducted two separate cluster analyses, one of colleges ($n = 15$), the other of departments ($n = 30$). Student grades were converted into numeric values (F = 0, D = 1, C-1.7, C = 2.0, C + =2.3, B- = 2.7, B = 3.0, B + =3.3, A- = 3.7, A = 4.0). We then computed and standardized the mean GPA based on the individually assigned grades across colleges and departments as the clustering variable for each sample.

We based our choice of clustering procedure on recommendations by Garson (2014) and others (Hahs-Vaughn, 2016; Mooi & Sarstedt, 2011). We used a two-step procedure that employed both hierarchical (i.e., within-groups linkage) and partitioning (i.e., *k*-means) methods, which address problems brought about by using either method on its own. We used nonhierarchical methods to identify empirically the number of homogenous clusters in the data. We then validated the stability of the cluster solution for each of the two samples using a hierarchical clustering method. We used a combination of empirical and rational methods to select the cluster solution with the appropriate number of clusters.

Finally, to further explore patterns in mean GPA, as well as the distributional characteristics of grades associated with the patterns identified in the *t*-test and cluster analyses, we computed the interquartile range (IQR), based on mean GPA, along with the corresponding variance, skewness, and kurtosis for each college/department in order to provide a more in-depth analysis of how grades might be internally differentiated. We took these steps as additional inductive means to uncover patterns based on mean GPA.

Results

Table 2 shows the descriptive statistics for the mean GPA of grades across the entire university, as well as for the samples of colleges and departments. Consistent with the findings of previous research (Rojstaczer & Healy, 2012), the majority of university grades were A's and B's. The mean GPA for the university was 3.34 ($SD = .82$); 86.7% of grades were A's and B's (F's comprised less than 2% of all grades). The mean GPA across the sample of colleges was 3.45 ($SD = .21$), across the arts and sciences departments, 3.78 ($SD = .24$). These findings are consistent with overall grade trends discussed at the outset. However, while grades are arguably high within the university as a whole, important questions remain about mean GPA by colleges and departments. In the following sections we explore this issue empirically.

t-Test

Using a stratified random sample of grades from the data ($n = 46{,}938$), we examined two subgroups within the university for internal differentiation in grade allocation. To address propositions 2a and 2b, we compared the mean GPA between departments in the social sciences, arts, and humanities and departments in the physical and biological sciences and mathematics. Consistent with propositions 2a and 2b, the independent samples *t*-test revealed that mean GPA in departments of social sciences, arts, and humanities ($M = 3.44$, $SD = .72$, $N = 28{,}186$) was significantly higher than those in departments of physical and biological sciences and mathematics ($M = 3.07$, $SD = .96$=, $N = 18{,}752$), $t(46{,}938) = 47.75$, $p < .001$, two-

Table 2 Descriptive statistics for Mean GPA across University, Colleges, and departments

Mean GPA	N	Min.	Max.	Mean	SD
Mean GPA University	460,562	0.00	4.00	3.34	.82
Mean GPA colleges	15	3.13	3.89	3.45	.21
Mean GPA Arts & sciences	31	2.82	3.78	3.31	.24
Mean GPA Colleges standardized	15	−1.52	2.14	0.00	1.00
Mean GPA Arts & sciences standardized	31	−2.03	1.93	0.00	1.00

tailed. This constitutes a body of evidence of variation by field consensus as outlined in the propositions. Grades are higher on average in the humanities and social sciences than elsewhere in the arts and sciences fields. While we thus have preliminary evidence of internal differentiation in the university grading structure, we investigated further differential grade distribution using cluster analysis techniques.

Cluster Analysis

The results of the cluster analyses of mean GPA provided a more robust descriptive analysis of how colleges and departments cluster within the university by the average grade allocated to students. To ease the interpretation of clusters, the mean GPAs for each college and department were standardized before entry into the cluster analysis (Huberty, Jordan, & Brandt, 2005). Following hierarchical clustering to identify empirically the number of homogenous clusters in the data, inspection of cluster analysis diagnostics enabled us to determine the number of clusters to retain from the data. We determined that a two-cluster solution was appropriate for the sample of colleges and a three-cluster solution for the sample of arts and sciences departments. We used *k*-means clustering to validate the stability of the cluster solution. Strongly supported by the data, the cluster solutions were consistent with the theoretical expectations. Table 3 illustrates the final cluster membership of colleges and departments, as well as the final standardized cluster center for each group.

The data reveal that colleges clustered around two centers—.5 standard deviations below the mean and 1.34 standard deviations above the mean—representing clusters of low and high mean GPA, respectively (Table 4). In the analysis of colleges, the *k*-means analysis yielded 11 departments (73%) clustered around a lower standardized mean GPA ($M = -.49$) and 4 departments (27%) clustered around a higher standardized mean GPA ($M = 1.34$). Consistent with proposition 1a, there is evidence of significant internal differentiation in grade allocation. As it turns out, however, the college-level data, like institutional-level data, likely mask important variation by field. It is difficult to ascertain patterns by field consensus even with

Table 3 Final cluster membership and centers for Colleges/Departments

Cluster membership	Colleges	Final cluster Centers (Std.)
Cluster 1: Low standardized mean GPA	11	−.49
Cluster 2: High standardized mean GPA	4	1.34
Arts & Sciences		
Cluster 1: Low standardized mean GPA	11	−1.03
Cluster 2: Avg. standardized mean GPA	13	.20
Cluster 3: High standardized mean GPA	6	1.51

college-level data. Thus we must examine departments to make more reliable inferences about patterns between consensus and grades.

Consistent with proposition 1b, we found that grades vary across departments. The cluster analysis showed that groups of departments within the arts and sciences clustered around three centers—1.03 standard deviations below the mean, .2 standard deviations above the mean, and 1.51 standard deviations above the mean—representing clusters of low, average, and high mean GPA, respectively. The *k*-means analysis yielded 11 departments (37%) clustered around low standardized mean GPA ($M = -1.03$), 13 departments (43%) clustered around the average mean GPA ($M = .20$), and 6 departments (20%) clustered around a high mean GPA ($M = 1.51$) (Table 4).

At the department level, patterns in grade distribution generally conform to postulated expectations with field consensus. Humanities fields concentrate in the high-GPA cluster, whereas natural science and mathematics departments concentrate in the low-GPA cluster (Table 4). There are high and low clusters, but also a middle cluster, which, too, is suggestive of pockets of variation in grade distribution and possible rigor. Cluster two gravitates slightly more toward cluster three (difference in standard deviation of 1.31) than toward cluster one (difference of 1.23). This tilt may be brought about by a preponderance in cluster two of several humanities and social science fields (e.g., anthropology, philosophy, theatre and film studies). (These gravitational pulls are captured further in Table 6, discussed below.) The department-level data are highly suggestive of a grading architecture consistent with expectations about field consensus.

Table 4 Cluster membership for colleges and departments

Colleges		
Cluster 1:Low GPA ($M = -.49$)	Cluster 2: High GPA ($M = 1.34$)	
College of Arts & Sciences	College of Education	
College of Business	School of Social Work	
College of Engineering	College of Public Health	
School of Pub and Intl Aff	College of Pharmacy	
College of Veterinary Medicine		
College of Env and Design		
College of Agr and Env Science		
School of Ecology		
College of Jour and Mass Comm		
College of Fam and Consumer Sci		
Departments		
Cluster 1:Low GPA ($M = -1.03$)	Cluster 2: Avg. GPA ($M = .20$)	Cluster 3: High GPA ($M = 1.51$)
Genetics	Geology	Poultry Science
Physics and Astronomy	Biological Sciences	Comparative Literature
Biochem and Molecular Bio	English	School of Art
Microbiology	Marine Sciences	Religion
Computer Science	Anthropology	Dance
Cellular Biology	Sociology	School of Music
Mathematics	Geography	
Statistics	Philosophy	
History	Psychology	
Chemistry	Romance Languages	
Communication Studies	Theatre and Film Studies	
	Classics	
	Germanic and Slavic Studies	

Interquartile Range

Finally, to assess propositions 1a-1c more fully, we explored the mean GPA of grades as well as the distributional characteristics of grades in the sample of colleges ($N = 15$) and departments ($N = 30$). We calculated the interquartile range (IQR) based on mean GPA for each college and department, as well as the corresponding variance, skewness, and kurtosis for each college/department in order to provide a deeper descriptive examination of the distributional characteristics of grades assigned across each subgroup. Table 5 shows the descriptive statistics for the mean GPA of each college, as well as the distributional characteristics associated with each college.

The lowest mean GPA ($M = 3.13$) was found in the college of business, while the highest ($M = 3.89$) was found in the school of social work. The highest variation in the distribution of grades was found in the college of veterinary medicine ($\sigma = .83$) and the lowest variance in grade allocation in the school of social work ($\sigma = .12$). Table 6 shows the descriptive statistics for the mean GPA for each department in the arts and sciences, as well as the distributional characteristics of grade allocation for each department. The lowest average departmental GPA was found in mathematics ($M = 2.82$), the highest in dance ($M = 3.78$). Likewise, the highest variance in grade allocation was found in mathematics ($\sigma = 1.22$), the lowest in dance ($\sigma = .22$). Once again, the patterns generally conform to expectations about field consensus. Humanities and social sciences departments concentrate toward the high end of mean GPA, natural sciences departments toward the low end.

Examining Tables 5 and 6, a clear pattern emerges in the IQR of mean GPA from top to bottom, along with the corresponding grade distributions for each college and department. As GPA increases across colleges and departments, the distribution of grades is squeezed into a tighter (leptokurtic) and higher end (left-skewed) of the grading distribution. Colleges and departments in the 25th percentile, with the lowest average GPAs, generally had higher variance in grade allocation. At the other end of the spectrum we found that colleges and departments in the 75th percentile, with higher average GPAs, had the lowest variance in grade distribution. In other words, for colleges and departments in the 75th percentile, not only are

Table 5 Mean GPA and distribution of grades for colleges

Department		Mean GPA	Variance	Skewness	Kurtosis
College of Business	25th Percentile	3.13	0.72	−1.20	4.60
College of Vet Medicine		3.22	0.83	−1.34	4.51
College of Engineering		3.23	0.80	−1.47	5.21
College of Arts and Sciences	50th Percentile	3.30	0.71	−1.69	6.28
College of Env and Design		3.31	0.66	−1.55	5.61
School of Forestry		3.37	0.55	−1.44	5.40
School of Pub and Intl Aff		3.35	0.59	−1.87	7.60
College of Fam and Consumr Sci		3.44	0.61	−2.03	7.99
College of Agr and Env Science		3.46	0.56	−1.77	6.84
School of Ecology		3.47	0.53	−1.84	7.29
College of Jour and Mass Comm	75th Percentile	3.50	0.38	−1.84	7.99
College of Public Health		3.64	0.38	−2.35	10.07
College of Pharmacy		3.66	0.44	−2.34	8.58
College of Education		3.70	0.31	−2.88	14.39
School of Social Work		3.89	0.12	−5.40	43.65

Table 6 Mean GPA and grade distribution for departments

Department		Mean GPA	Variance	Skewness	Kurtosis
Mathematics	25th Percentile	2.82	1.22	−0.99	3.35
Chemistry		2.92	1.01	−1.13	4.01
Genetics		3.01	0.99	−1.09	3.83
Microbiology		3.04	0.75	−1.03	4.14
Physics and Astronomy		3.11	0.79	−1.17	4.42
Biochem and Molecular Bio		3.11	0.78	−0.99	3.75
Statistics		3.11	0.94	−1.29	4.35
History		3.12	0.64	−1.58	6.54
Computer Science	50th Percentile	3.12	1.01	−1.49	4.97
Communication Studies		3.18	0.49	−1.43	6.54
Cellular Biology		3.18	0.64	−1.19	4.79
Sociology		3.24	0.63	−1.41	5.58
Plant Biology		3.25	0.81	−1.63	5.84
Philosophy		3.29	0.68	−2.07	8.12
Germanic and Slavic Studies		3.29	0.70	−1.67	6.26
Biological Sciences		3.29	0.70	−1.65	6.10
Marine Sciences		3.31	0.57	−1.70	7.02
Psychology		3.34	0.62	−1.70	6.67
Geography		3.35	0.60	−1.68	6.62
Geology		3.38	0.59	−1.59	6.08
Classics		3.40	0.73	−2.02	7.52
Anthropology		3.40	0.59	−2.06	8.29
English	75th Percentile	3.46	0.40	−2.07	10.34
Romance Languages		3.46	0.46	−2.03	8.88
Theatre and Film Studies		3.52	0.53	−2.20	9.07
Poultry Sciences		3.55	0.44	−1.84	6.94
School of Art		3.62	0.37	−2.61	12.52
Comparative Literature		3.68	0.40	−3.24	16.05
Religion		3.70	0.37	−3.22	16.13
School of Music		3.75	0.31	−3.23	16.38
Dance		3.78	0.22	−4.51	31.63

grades primarily located in the upper end of the grading distribution for colleges and departments, they are also the most tightly compressed.

For example, colleges with a lower mean GPA, such as business, veterinary medicine, and engineering, also have the highest variance in grade allocation, as well as lower skewness and kurtosis of grade distributions. The college of business has a mean GPA of 3.13, variance of .72, skewness of −1.2, and kurtosis of 4.6, whereas the school of social work has the highest mean GPA (3.89), a variance of .12, skewness of −5.40, and kurtosis of 43.65. Thus, grades are compressed toward the upper end of the distribution in the school of social work as compared to the college of business.

The same patterns are found when examining the IQR and distributional characteristics of departments within the college of arts and sciences (Table 6). Within the college of arts and sciences, mathematics and chemistry have the lowest mean GPA and very low skewness and kurtosis, while also the highest variance in grades. By contrast, religion, music, and dance have the highest mean GPA, along with highly skewed and kurtotic grading distributions.

Mathematics, the department with the lowest GPA (2.82), has a variance of 1.22, a skewness of −0.99, and kurtosis of 3.35, in comparison to dance, which has the highest mean GPA (3.78), the most left-skewed distribution (−4.51), and the most leptokurtic (compressed) distribution (31.63). Thus, we find that in departments with the highest GPA (e.g., art, comparative literature, religion, music, and dance), not only are grades primarily located in the upper end of the grading distribution, they are the most tightly compressed in the upper end of the distribution. Put differently, not only do these fields have high GPAs, but very large proportions of students have these GPAs. There is little dispersion. These findings correspond to fields characterized by low consensus.

We have found that the distributions of grades for colleges and departments in the 75th percentile of mean GPA were less symmetrical, had heavier upper tails, and lower variance. By contrast, colleges and departments in the 25th percentile of mean GPA had grade distributions that were more symmetrical, lighter upper tails, and higher variance in grade allocation. These patterns support propositions 1a-1c: both assigned grades and grade distributions varied significantly across the university. Examination of the IQR and distributions of grades associated with colleges and departments provide ample evidence of variation in the distribution of grades across fields in the university.

Discussion and Conclusion

Substantial research has documented grade inflation in American universities, but extant research has not examined patterns in how grades might be more variously distributed among fields. This study has documented significant variation of grade patterns *inside* a university. Colleges and departments clustered around different groups of geometric difference in the standardized mean GPA. For colleges, two main clusters emerged around low and high average GPA. For departments, three cluster centers—representing low, medium, and high mean GPA—were identified.

Fields that clustered around higher standardized mean GPAs, along with their low variance and highly left-skewed and kurtotic grade distributions, generally correspond to fields of low consensus. By contrast, fields that clustered around lower average GPAs, along with their higher variance in grade distributions and lower skewness/kurtosis, generally correspond to higher consensus fields. Not all fields are decidedly at high or low ends on a continuum of consensus; they are in between. To that end, medium consensus groups had z-scores that clustered in between the relatively high and low consensus groups. The cluster analysis identified colleges, such as arts and sciences, business, and engineering, that—*in aggregate*—may represent comparatively high consensus fields (i.e., they had mean GPAs that were substantively dissimilar to other colleges in the university).

Colleges that clustered with comparatively high GPAs may do so by two different processes: one, by the allocation of higher grades (e.g., education and social work in the analysis); and, two, by selection processes that recruit especially high achieving students (e.g., public health and pharmacy in the analysis). A limitation is that we could not control for prior academic achievement of students, characteristics of instructors, or characteristics of the organizational structure in this analysis. Therefore, it remains unknown whether or how much the differences in grade variation might be attributable to student-, instructor-, or organizational- factors. Future research should examine the extent to which variation in grade allocation is attributable to student-specific, instructor-specific, and/or university-specific characteristics.

At the department level, mathematics, genetics, and chemistry are characterized by substantively different grade allocation. These programs, generally high consensus fields, were also characterized by the highest variance in grades. At the other end of the spectrum, fields such as comparative literature, art, religion, dance, and music—generally fields of lower consensus—clustered around lower mean GPAs and evinced the lowest variance in grades.

In their book *Academically adrift*, Arum and Roksa (2011) followed over 2,300 students at 24 universities throughout the U.S. for four years to find that learning is limited on college campuses. According to these authors, a large number of students showed no significant progress on tests of critical thinking that were administered when they began college and then again at the end of their sophomore and senior years. Arum and Roksa argued that many college courses and majors are not rigorous; and yet, they observed, college grades are high.

We do not assume that high grades are explained by inflation exclusively. For example, Mostrom and Blumberg (2012) found evidence that students learn more by improved teaching, which can result in higher overall grades. At the same time, we question whether grades, as pervasively high as they are, are explainable by improved teaching and learning alone, especially in light of coincident trends that document significant decreases in the amount of time that undergraduates spend studying as well as ways that "academic capitalism" may impinge upon students in U.S. universities, which we discuss below.

Babcock and Marks (2011) documented that the average number of hours per week college students spent on academics in the 1960s was approximately fifteen hours in class and twenty-five hours studying, whereas today it is fifteen hours in class and twelve to thirteen hours studying. Studies of college students' time use show that they place an emphasis on the social realm where sociability and sensitivity to social groups are highly valued (Arum, Roksa, Cruz, & Silver, 2018). Brint and Cantwell (2010) found that each week students spent thirteen hours studying, fourteen hours working, seventeen hours watching television, and twenty-four hours engaged in other forms of entertainment and socializing. Thus students spent over forty hours each week in leisure and social activities—over three times the amount of time they spent studying (see also Charles, Fischer, Mooney, & Massey, 2009).

Patterns of high grades also coincide with an ascendance of "academic capitalism"—the increasing corporatization of universities—the consequences of which reach students. Slaughter and Rhoades (2004) argued that universities increasingly engage students as "consumers" to whom "the college experience" is marketed as a time in "attractive places in which to live, consume services, and play [rather] than as challenging places in which to learn and become educated (Slaughter & Rhoades, 2004, p. 298). Likewise arguing that students have adopted a market orientation to college and to course-taking, Grigsby contended that an attitude of earning a degree with as little effort as possible has emerged (Grigsby, 2009).

While grade inflation has been found to be widespread across universities in the U.S., we suggest that universities are not monolithic organizations in how grades are distributed. There is, rather, a complex architecture in the allocation of grades across colleges and departments. Structured differentials in grade allocation may indicate variation in rigor and learning, an important issue not explored by research on grade inflation and related work on rigor (e.g., Arum & Roksa, 2011). We believe that our findings raise questions about the widely accepted view that student learning in postsecondary education in the United States is uniformly low. If grades are taken as a measure of rigor, universities are internally varied across fields where rigor and associated learning are found. For this study we used the concept of field consensus as a means to differentiate the grade structure of universities. The data support the postulate that it is more difficult to achieve high grades in fields marked by clarity and consistency of standards. By contrast, it is easier to achieve high grades in fields in which the standards are less clearly defined.

We suggest that our findings reveal a need to examine further the linkages between field consensus, rigor, student learning, and grade allocation in the contemporary university. Examining grades *only at an institutional level*, which has been the customary approach, masks a highly significant variation in grade distributions inside universities. It is time that we now examine more thoroughly "what goes into" a course grade and how grading practices vary across fields in a university.

Acknowledgments The authors thank Karlo Lei and Sondra Barringer for comments and suggestions in the development of the paper.

References

American Association of University Professors (2015). *Policy documents and reports* (11th ed.). Baltimore, MD: Johns Hopkins University Press.

Arum, R., & Roksa, J. (2011). *Academically adrift: Limited learning on college campuses*. Chicago, IL: University of Chicago Press.

Arum, R., Roksa, J., Cruz, J., & Silver, B. (2018). Student experiences in college. In B. Schneider (Ed.), *Handbook of the sociology of education in the 21st century* (pp. 385–403). Dordrecht, The Netherlands: Springer.

Babcock, P. S., & Marks, M. (2011). The falling time cost of college: Evidence from half a century of time use data. *Review of Economics and Statistics, 93*, 468–478.

Becher, T. (1989). *Academic tribes and territories: Intellectual enquiry and the cultures of disciplines*. Milton Keynes, United Kingdom: Open University Press.

Becker, H.S., Geer, B., & Hughes, E.C. (1968). *Making the grade: The academic side of college life*. New Brunswick, NJ: Transaction.

Braxton, J. M., & Hargens, L. L. (1996). Variation among academic disciplines: Analytical frameworks and research. In J. C. Smart (Ed.), *Handbook of theory and research* (Vol. 11, pp. 1–46). New York, NY: Agathon.

Brint, S., & Cantwell, A. M. (2010). Undergraduate time use and academic outcomes: Results from the University of California undergraduate experience survey, 2006. *Teachers College Record, 112*, 2441–2470.

Charles, C. Z., Fischer, M. J., Mooney, M. A., & Massey, D. S. (2009). *Taming the river: Negotiating the academic, financial, and social currents in selective colleges and universities*. Princeton, NJ: Princeton University Press.

Cole, S. (1993). *Making science: Between nature and society.* Cambridge, MA: Harvard University Press.

DeMarzo, P. M., Fishman, M. J., & Hagerty, K. M. (2005). Self-regulation and government oversight. *Review of Economic Studies, 72*, 687–706.

Frey, B. S. (2006). Giving and receiving awards. *Perspectives on Psychological Science, 1*, 377–388.

Garson, G. D. (2014). *Cluster analysis.* Ashboro, NC: Statistical Associates.

Goldman, L. (1985). The betrayal of the gatekeepers. *The Journal of General Education, 37*, 97–121.

Grigsby, M. (2009). *College life through the eyes of students*. Albany, NY: State University of New York Press.

Hagstrom, W. O. (1965). *The scientific community.* New York, NY: Basic Books.

Hahs-Vaughn, D. (2016). *Applied multivariate statistical concepts*. New York, NY: Routledge.

Hargens, L. L. (1975). *Patterns of scientific research: A comparative analysis of research in three scientific fields*. Washington, DC: American Sociological Association.

Hargens, L. L., & Kelly-Wilson, L. (1994). Determinants of disciplinary discontent. *Social Forces, 72*, 1177–1195.

Hermanowicz, J. C. (1998). *The stars are not enough: Scientists—their passions and professions*. Chicago, IL: University of Chicago Press.

Hermanowicz, J. C. (2009). In *Lives in science: How institutions affect academic careers*. Chicago, IL: University of Chicago Press.

Hu, S. (2005). Beyond grade inflation: Grading problems in higher education. ASHE Higher Education Report, 30. San Francisco, CA: Jossey-Bass.

Huberty, C.J., Jordan, E.M., & Brandt, W.C. (2005). Cluster analysis in higher education research. In J.C. Smart (Ed.), *Higher education: Handbook of theory and research, Vol. 20* (pp. 437–457). Dordrecht: The Netherlands: Springer.

Hunt, L.H. (Ed.). (2008). *Grade inflation: Academic standards in higher education*. Albany, NY: State University of New York Press.

Johnson, V. E. (2003). *Grade inflation: A crisis in college education*. New York, NY: Springer.

Krautmann, A. C., & Sander, W. (1999). Grades and student evaluations of teachers. *Economics of Education Review, 18*, 59–63.

Merton, R. K., & Zuckerman, H. ([1972] 1973). Age, aging, and age structure in science." In N. W. Storer (Ed.), *The sociology of science: Theoretical and empirical investigations* (pp. 497–559). Chicago, IL: University of Chicago Press.

Milton, O., Pollio, H. R., & Eison, J. A. (1986). *Making sense of college grades*. San Francisco, CA: Jossey-Bass.

Mooi, E., & Sarstedt, M. (2011). *A concise guide to market research*. Berlin, Germany: Springer.

Mostrom, A. M., & Blumberg, P. (2012). Does learning-centered teaching promote grade improvement? *Innovative Higher Education, 37*, 397–405.

Pascarella, E. T., & Terenzini, P. T. (2005). *How college affects students, volume 2*. San Francisco, CA: Jossey-Bass.

Rojstaczer, S. (2016). *GradeInflation.com: Grade inflation at American colleges and universities*. Retrieved from http://www.gradeinflation.com/

Rojstaczer, S., & Healy, C. (2010). Grading in American colleges and universities. *Teachers College Record*. http://gradeinflation.com.tcr2010grading.pdf

Rojstaczer, S., & Healy, C. (2012). Where a is ordinary: The evolution of American college and university grading, 1940-2009. *Teachers College Record, 114*, 1–23.

Rosovsky, H., & Hartley, M. (2002). *Evaluation and the academy: Are we doing the right thing?: Grade inflation and letters of recommendation*. Cambridge, MA: American Academy of Arts & Sciences.

Slaughter, S., & Rhoades, G. (2004). *Academic capitalism and the new economy: Markets, state, and higher education*. Baltimore, MD: Johns Hopkins University Press.