

Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Copyright is retained by the first or sole author, who grants right of first publication to the *Practical Assessment, Research & Evaluation*. Permission is granted to distribute this article for nonprofit, educational purposes if it is copied in its entirety and the journal is credited.

Volume 15, Number 14, October, 2010

ISSN 1531-7714

Grades and Ranking: When Tenure Affects Assessment

Jean Filetti, Mary Wright, William M. King
Christopher Newport University

This article examines how a faculty member's status—either tenured or tenure-track—might affect the grades assigned to students in a writing class. We begin with a brief review of the research surrounding faculty to student assessment practices and follow with specific controversies regarding faculty motivation pertaining to grading practices. We interpret the grade distributions of tenured and tenure-track faculty members teaching a sophomore-level writing course in an English Department at a small, public liberal arts university in Virginia, examine the relationship between grade distributions and faculty rank, and conclude that tenure-track faculty grade more leniently than their tenured colleagues, primarily in the number of “A” grades assigned. The results of this study suggest that tenured professors tend to award fewer “As” than tenure-track professors. We posit that at universities where emphasis is placed upon teaching, two specific patterns emerge: reciprocity may be an unspoken agreement between some faculty and students with regard to the exchange of good grades for good evaluations, or with experience comes rigor.

Issues surrounding grading have long plagued faculty and students in every department of every college and university. Students believe they work hard for their grades and take the results personally, as evidenced in their verb choices within remarks, such as, “She *gave* me a ‘C,’” instead of “I *earned* a ‘C,’” and faculty feel they fairly assess the results with the correspondingly deserved mark. Many departments require faculty members to disclose their grading rubrics on syllabi and offer “norming” or calibrating sessions, workshops where attendees, mostly tenure stream and restricted faculty, assess sample student papers and attempt to form a consensus on evaluating sample assignments. These faculty members are motivated by the knowledge that a good deal depends upon what constitutes a grade and that department chairs to the president's office track faculty grade point averages (GPAs). In yearly evaluations faculty members must rationalize their processes, explain anomalies, and otherwise justify this very personal business of assessing student work.

This article examines whether the faculty member's status—either tenured or tenure-track—affects the

grades assigned to students in a writing class. The initial hypothesis was that tenure-track faculty were more likely to grade leniently because (1) their status at the university was not secured by tenure and (2) because they desired to receive positive student evaluations on their teaching (teaching is often heavily weighted in administrative evaluations of faculty) and believed that awarding higher grades would translate into higher student evaluations.

In assembling the data for this study, we examined the grade distributions of tenured and tenure-track faculty members in a sophomore-level writing seminar. We begin with a brief review of the research surrounding faculty to student assessment practices and follow with specific controversies regarding faculty motivation regarding grading practices. We interpret data for tenured and tenure-track faculty members in an English Department who teach a sophomore-level, topics-based research course at a small, public liberal arts university in Virginia, raising the issue regarding how tenure identity contributes to evaluation decisions. We present our findings and examine the relationship between course GPAs, tenured, and tenure-stream faculty and conclude

that tenure-track faculty grade more leniently than their tenured colleagues primarily in the number of “A” grades assigned. The evidence suggests there exists a triad between faculty identity, student reactions, and assessment in higher education and may be tied to the exchanges of power and agency between administration, faculty, and students.

RELATED RESEARCH

As studies from the last thirty years demonstrate, one central and unchanging combination ranks highest on the list of mitigating reasons behind grading habits and strategies: student evaluations of teaching (SETS) and the subsequent ways institutions use the reports (Coladarci & Kornfeld, 2007). Others have used this pairing to canvas the mitigating factors that affect grade decisions, though few consider faculty ranking a variable, as our results suggest. Fred Pestello (1987) examines how organizational setting in a course with multiple sections affects grades to explain how one faculty member, Dr. Goodgrade (tenure status unknown), manipulates his graduate teaching assistants’ grading practices to ensure an extraordinary small variation between their final results. Pestello concludes that Goodgrade’s efforts result from his desire to boost sagging enrollments, prevent administrators from reacting to grading extremes, and protect his TAs from becoming either negatively notorious or too popular because of their evaluation practices. Since he was one of Goodgrade’s TAs, Pestello includes a warning against allowing such institutional concerns to eclipse the obvious human presence in grading processes.

Another central issue in evaluation practices centers upon how the systematic ritual of grading affects relationships between faculty and their students. Zangenehzadeh (1988) cites grade inflation as a direct result of student evaluations, but maintains that researchers may not use final grades for an accurate reflection since SETS are administered before the submission of final papers or other instruments. He considered how students with higher GPAs may evaluate their professors more positively and how attitudes towards the course and the instructor affect SETS outcomes before suggesting that, similar to Pestello’s Dr. Goodgrade study, the kind of norming that occurs when individual faculty scores are adjusted to the school’s or nation’s standards might help faculty aspire to results higher than average if they know lower ratings will not be considered acceptable in merit evaluations.

After breaking down his statistical evidence and yielding a variety of results, Eiszler (2002) concludes that faculty members either become more rigorous graders by cooperative and or instructional materials, or with age and experience. Likewise, Hamilton’s data (1980) supports Eiszler’s. Hamilton finds newer faculty have a more difficult time currying favor with their students than their experienced colleagues, though he categorizes the inexperienced as part-time and experienced as full time without further distinction between tenured and untenured. Not surprisingly, Hamilton adds rapport between students and professors as an additional element for consideration.

Nasser and Hagtvet (2006) claim that data corresponding to an aggregate, such as the faculty members as individuals and as a mass, render faulty findings because a single number, such as a course GPA, yields little regarding them as individual students and their GPAs for all courses. Therefore, the collective course GPA may have little resemblance to either a collective course GPA in, for example, a 200-level history or psychology course, or to students’ GPAs for other courses they took concurrent to the 223 class in this study. Nasser and Hagtvet’s findings reveal an overwhelming correlation between the expected final grade and positive instructor evaluation. Students who believed they would receive a higher grade responded more favorably than those who felt they would get a lower grade. Moreover, final grade expectations, the pair concludes, cannot be associated with grading leniency or effective teaching, ideas posited elsewhere (Marsh, 1987, Young and Shaw, 1999). Interestingly, Nasser and Hagtvet suggest a more effective procedure to solicit data would involve a beginning evaluation as well as the ending instrument in order to measure more consistent movement in student evaluation and to observe important trends from faculty and students. Regarding faculty status, they conclude that data interpretation remains key to determining whether results are legitimate or biased, meaning the reduction of bias (perhaps concealing the professor’s rank) should reduce the different data.

In their study conducted at a small private college using data from the business school collected over a 20-year period, Kezim, Pariseau and Quinn (2005) conclude that faculty ranking bears a direct and obvious impact on grade inflation, citing differences between tenured and adjunct faculty and non-tenured and adjunct faculty in order to support their claim that the adjunct faculty’s consistently higher grades contribute to

institutional grade inflation. However, although they conclude that significant differences occurred between adjuncts and tenured and adjuncts and non-tenured, the differences between tenured and non-tenured were nominal at best, thus calling into question Moore and Trahan's (1998) conclusion that non-tenured faculty attempt to "buy" tenure by inflating grades. Interestingly, in charting the individuals and their arcs to tenure, the three observe how the first nontenured to achieve tenure graded more stringently than the last. This supports their (and many others') hypothesis that grade inflation has steadily risen in higher education within the last 30 years.

Stratton, Myers, and King (1994) suggest that the faculty member's experience, education, and pedagogical attitudes affect student SETS responses, in addition to grades, in their study conducted at the University of Akron from 1981-1990. Of considerable note that speaks to faculty motivation is the fact that, in the Economics Department, for example, 45 percent of faculty pay is determined by teacher effectiveness, as revealed by SETS narratives and scores, and retention and tenure are measured by the same information. Stratton, Myers, and King try to determine whether faculty members deliberately adjust their behaviors to rate higher with their students, sampling 10 faculty members who taught before SETS were administered and after. They find that grades increase in value with the instructor's experience, a result inconsistent with all other studies represented herein, and that the cohort's grades briefly spiked to an 11 percent increase with the introduction of the SETS, but over time this increase dissipated. To wit, they warn administrators to consider the relationship between student grades and SETS scores may be the result of events or phenomena outside of or preexisting to the evaluation instrument.

METHODOLOGY

For the purposes of examining grade distributions in classes taught by tenured faculty within the English Department compared to grade distributions in classes taught by tenure-track faculty in the English Department, the study selected a second-year writing seminar (hereafter called English 223).

In the 2004-2005 academic year, the English Department redesigned its Freshman Writing Program (a two-course freshman writing sequence positioned in the student's first year) and created a Writing Program (a two-course writing sequence redistributing the second required composition course to the student's

sophomore year). The newly created sophomore composition course places students in a collaborative research community formed around a stimulating topic selected by the individual instructor. It builds upon skills covered in the freshman writing seminar (analyzing and producing several short but sophisticated written arguments within ongoing social and cultural questions), but integrates the conventions of academic research and guides students through staged assignments towards a final, polished 10-15 page research paper.

Taught for the first time in the 2004-2005 academic year, English 223 offers the unique opportunity to examine grade distributions in a course new to both tenured and tenure-track faculty. In other words, faculty members were on a relatively even playing field. Along with their tenure-track colleagues in the department, senior tenured faculty members were also teaching the course for the first time. Therefore, previous experience teaching this particular course was not a factor.

Additionally, the course requirements and the Writing Program's training workshops provide a degree of consistency to English 223. For example, the two cohorts (tenured and tenure-track) teaching this course attend workshops on assignment construction, staged writing and research activities, effective conferencing, evaluating writing using rubrics, and other pedagogical issues pertaining to the best practices in academic research and writing. Instructors are also required to integrate the following writing requirements: (1) informal writing (notes, journals, early drafts, and posts to online prompts), (2) semi-formal writing (mid-drafts, in-class essays, and reading responses), and (3) formal writing (a target of 15 pages of polished writing that integrates multiple sources to be evaluated on content, correctness, organization, and style). Finally, faculty members use *The Brief Wadsworth Handbook* and its companion Web site (InSite) to govern the writing/research process.

Finally, all sections of this sophomore-level writing seminar have an enrollment cap of 19, with the classes comprising the data base consisting of 15 to 19 students. The faculty members teaching this course, therefore, have similar paper loads and instructional time for individualized instruction. Since all faculty at this university teach a 4-course load per semester (unless they receive release time for research or administrative duties), both junior tenure-track and senior tenured professors have similar teaching responsibilities.

Student Demographics

All students enrolled in English 223 must have sophomore standing (completion of 30 credit hours) and have received a “C-“ or better in English 123. Unlike freshmen who populate the first composition course (English 123), students in English 223 represent a demographic population that has largely adjusted to college-level academic expectations and social transitioning. Furthermore, unlike English 123, from which students with high Advanced Placement (AP) scores or International Baccalaureate (IB) course work, can exempt, all students must take English 223 (unless they are transfer students who have earned an Associates Degree or have taken a course similar to English 223 at another institution). In other words, English 223 is the only course in this university’s curriculum that all students admitted to this university as freshmen must take.

Report Data

We examined the grade distribution reports for all 33 sections of English 223 taught within the English Department for two academic years or four semesters. Sixteen of those sections were taught by tenured English faculty and 17 were taught by tenure-track English faculty. Those students who completed the course and received letter grades totaled 595.

Although some students may have been retaking the course in semesters 2, 3, and 4 because they did not pass it with a “C-“ or higher during a previous semester, they would have been spread among the multiple sections offered, as there are no designated remedial sections. Furthermore, repeating students were not identifiable in the collected data, so there is no way of knowing if repeating the course resulted in a higher grade. Finally, because students enrolled themselves in their respective English 223 classes, the present study lacks random assignment and, thus, inhibits our ability to draw strong conclusions.

RESULTS AND ANALYSIS

Of the 16 classes (293 students) taught by tenured faculty (one full professor and six associate professors), the average class GPA was 2.87 with a standard deviation of 0.23. Of the 17 classes (302 students) taught by tenure-track faculty (five assistant professors), the average class GPA was 3.05 with a standard deviation of 0.38. To calculate class GPA, each letter grade administered in the English 223 class was assigned a numeric value. The values assigned to each grade are

as follows: an “A” receives a 4.0, “A-” receives a 3.7, “B+” receives a 3.3, “B” receives a 3.0, and so on until letter grade “F,” which receives 0.0. A two-sample *t*-test for average GPA revealed a nonsignificant difference between tenured and tenure-track faculty in grade point average for their respective English 223 classes, $t(31) = -1.64, p > .1$.

In order to determine whether the grades administered are independent from the tenure status of the faculty, we conducted the chi-square test of independence using an alpha level set at .05. The chi-square test of independence is a nonparametric measure of association between two categorical or ordinal variables. As letter grades lack numeric values that are equidistant from each other, this variable can be considered ordinal, and, thus, it is more appropriate to examine the differences between the two groups’ grade distributions using a nonparametric test. In order to avoid violating the test assumption of small cell sizes, grades for “D+,” “D,” and “D-” were collapsed into a single category, “Ds.” The chi-square test of independence revealed a significant association between the grades administered and the tenure status of the faculty, ($df=9, N = 595$) $X^2 = 24.85, p < .01$. The strength of the association, however, is relatively small, $C = .20, p = .003$. This Pearson’s Contingency Coefficient, C , is an effect size statistic and can range from 0 (very weak association) to 1 (very strong association). The significant association suggests that somewhere within the grade distributions there is a discrepancy between the grades we would expect if tenured and tenure-track faculty administered grades equally and the actual grades the faculty awarded their students.

To elucidate where the discrepancy lies, we computed standard residuals. According to Sheskin (1997), standard residuals with absolute values greater than 1.96 indicate that the cell accounts for significant contribution to the association between two categorical variables. Analysis of the standard residuals concluded that tenured and tenure-track faculty differed significantly in only one grade: “As” with -2.8 and 2.7 standard residuals for tenured and tenure-track faculty, respectively. Table 1 depicts the observed grades administered and their standard residuals for both tenured and tenure-track faculty.

For a graphical comparison of the actual grades awarded by tenured and tenure-track faculty see Figure 1.

Figure 1 highlights two important groupings of students—those who did very well in the course and those who failed the course, with tenure-track faculty awarding significantly more “As” or “A-s” to their students than tenured faculty. Results also show that the grades in classes taught by tenured faculty are moderately lower than grades in classes taught by untenured faculty.

Table 1: Observed Grade Frequency and Standard Residuals by Tenure Status

Grade Administered	Tenured		Tenure-Track	
	Observed	Std. Residual	Observed	Std. Residual
A	27	-2.8	66	2.7
A-	48	0.7	40	-0.7
B+	52	0.5	46	-0.5
B	68	1.0	55	-0.9
B-	31	-0.4	37	0.4
C+	20	0.8	14	-0.8
C	17	0.2	16	-0.2
C-	5	-1.0	11	1.0
Ds	7	0	7	0
F	18	1.1	10	-1.1

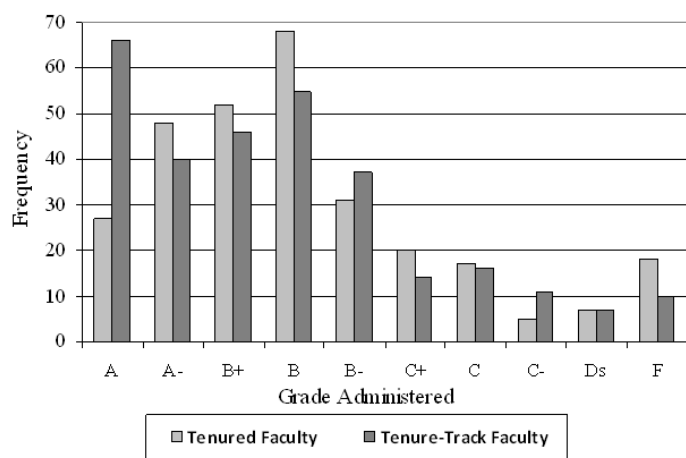


Figure 1 Grade Distribution by Tenure Status

LIMITATIONS

Certainly, multiple factors determine student performance and learning assessment. In this particular study, students interested in the course topic (global warming, political cartoons, piracy) may have been more fully engaged in researching and writing about the field and may, in fact, have produced better writing than

students with no or limited interest in the topic. That being said, however, given the fact that students often select a particular section of the second-year writing seminar based on other factors than interest in the topic (for example, the class meeting time fits their personal schedule or with other courses they are taking), it is likely that each section of English 223 was populated with students who self-selected the course based on the topic and students who selected the course based on other factors.

Furthermore, the study’s data are drawn from a small liberal arts university whose faculty members teach a heavy course load (4 courses per semester at the time of the study). Therefore, any conclusions regarding the relationship between rank and grading practices may not reflect that found at either research-centered institutions or larger liberal arts schools with reduced teaching loads and where evaluations for tenure may rest on a faculty member’s publication and research. Furthermore, the data are specific to a sophomore-level composition course; data from other disciplines or courses may not yield similar results. That being said, however, there is no other single course (other than English 123, the first-year writing seminar) required of undergraduate students at this university, and, as mentioned earlier, students may be exempted from the first-year writing seminar through AP credit and/or IB course work. Therefore, even if data were collected from courses in other disciplines, the students populating, for example, a general education math or government class would have self-selected that course.

CONCLUSIONS

The results of this study suggest that tenured faculty members at this small undergraduate liberal arts university grade differently from tenure-track faculty members. This difference is relatively small and restricted to only the highest grades. Specifically, tenured professors tend to award fewer “As” than tenure-track professors. The difference that exists between faculty cohorts could arise as a result of two possibilities: the students who made higher grades in courses taught by tenure-track faculty could have conceivably learned more than if taught by tenured faculty, or those tenure-track faculty members could have graded easier and their students may have learned less than if they were assessed more rigorously.

The results suggest that at universities where emphasis is placed upon teaching, reciprocity may be an unspoken agreement between some faculty and students

with regard to the exchange of good grades for good evaluations. By extension, we concur with those studies that find experience increases rigor, so tenured faculty members who have more years invested in their teaching careers may grade more stringently than they did as teaching assistants or newly hired tenure stream professors. Finally, in terms of administrative evaluation of faculty teaching, individual faculty grade distributions provide additional information to supplement that in student evaluations, especially in the absence of other methods of assessing teaching effectiveness, such as course administered pre- and post-test content exams.

REFERENCES

- Boualem K., Pariseau, S., & Quinn, F. (2005). Is grade inflation related to faculty status? *The Journal of Education for Business*, 80(6), 358-264.
- Coladarci, T. & Kornfield, I. (2007). RateMyProfessors.com versus formal in-class student evaluations of teaching: A study of concurrent validity. *Practical Assessment, Research & Evaluation*, 12(6).
- Eiszler, C. E (2002). College students' evaluations of teaching and grade inflation. *Research in Higher Education*, 43(4), 483-501.
- Hamilton, L.C. (1980). Grades, class size, and faculty status predict teacher evaluations. *Teaching Sociology*, 8(1), 47-62.
- Marsh, H.W. (1987). Students' evaluations of university teaching: Research findings methodological issues, and directions for future research. *International Journal of Educational Research*, 11(3), 263-388.
- Moore, M. & Trahan, R. (1998). Tenure status and grading practices. *Sociological Perspectives*, 41(4), 775-781.
- Nasser, F., & Hagtvet, K. (2006). Multilevel analysis of the effects of student and instructor/course characteristics on student ratings. *Research in Higher Education*, 47(5), 559-590.
- Pestello, F. (1987). The social construction of grades. *Teaching Sociology*, 15(4), 414-417.
- Sheskin, D.J. (1997). Test 11: The Chi-Square test for r x c tables. *Handbook of parametric and nonparametric procedures*. Boca Raton, FL: CRC Press.
- Stratton, R.W., Myers S. C., & King, R. (1994). Faculty behavior, grades, and student evaluations. *The Journal of Economic Education*, 25(1), 5-15.
- Young, S., & Shaw, G.D. (1999). Profiles of effective college and university teachers. *Journal of Higher Education*, 70(6), 670-686.
- Zangenehzadeh, H. (1988). Grade inflation: A way out. *The Journal of Economic Education*, 19(3), 217-226.

Citation:

Filetti, Jean, Wright, Mary & William M. King (2010). Grades and Ranking: When Tenure Affects Assessment. *Practical Assessment, Research & Evaluation*, 15(14). Available online: <http://pareonline.net/getvn.asp?v=15&n=14>.

Corresponding Author:

Jean Filetti, Chair
Department of English,
Christopher Newport University,
Newport News, VA 23606
Email: Filetti [at] cnu.edu