Course Difficulty and its Association with Student Perceptions of Teaching and Learning   *RESEARCH*

Amanda Joyce · Murray State University

Abstract

Grade inflation has long been an issue in academia, and with this comes the concern that instructors will feel pressured to inflate grades in order to improve student evaluations of their teaching. Many historical studies have demonstrated associations between higher grades and higher teaching evaluations. The purpose of this investigation was to determine the relationship between high grades and high teaching evaluations, and their association with other indicators of course difficulty. Anonymous, end-of-semester, teaching evaluations were collected from 156 students in 6 sections of 3 unique courses in the Psychology department of a large Southeastern University between 2011 and 2014. Students were asked to report on various aspects of their learning experience, including their instructor's effectiveness, the level of mutual respect in the classroom, and their expected grade in the course, among other variables. Students' agreement with the statement, "Overall, the instructor's teaching was effective," positively related to their evaluation of all individual aspects of the instructor's effectiveness (e.g., "The instructor was well-prepared"; "The instructor presented subject matter clearly"; all r's > .433; all p's < .001). However, student evaluations of overall instructor effectiveness showed no association with their expected grade in the course ($r = .133$, $p = .101$), nor with the number of writing assignments or exams given by the instructor (all r's < .138; all p's > .088). The results imply that instructors need not feel pressured to reduce course demands in order to improve student evaluations.

Keywords: grade inflation, course difficulty, instructor evaluations, student satisfaction

Grade inflation has long been of concern in academia, and with this comes the concern that instructors will feel pressure to inflate grades and to otherwise create easier courses in order to improve student evaluations of teaching. The purpose of this investigation was to determine if the association between high grades and high teaching evaluations exists today, as it has in the past, while also examining if individual indicators of course difficulty relate to students' perceptions of courses.

Concerns about grade inflation can be traced back to the 1970's (Bowers, 1970; Juola, 1976). Grade inflation occurs when student grades improve, but student achievement does not (Stone, 1995). Though students are happy to receive high grades in their courses, it is problematic to educators as well as to their future employers, because it compresses all grades at the top of a spectrum such that it is difficult to tell the best students from those who are only good, and the good students from those who are only mediocre (Johnson, 2006). Unfortunately, there is ample evidence of grade inflation during the late 20th century. One study investigating the impact of grade inflation from 1962 to 1985 found an increase in average grade point average from 2.49 to 2.93 (Sabot & Wakeman-Linn, 1991). Similarly, the percentage of students expecting an A or A- grade in a course increased by 10% across the 1990's (Eiszler, 2002).

These increasing grades are encouraged by a student populace who rewards universities and instructors for an artificially inflated grade point average. During the 1990's, as the number of students who expected an A or A- in a course increased, the average rating on student teaching evaluations also increased by .1 points (Eiszler, 2002). Indeed, other research supports the notion that students give more favorable course evaluations to instructors of easier courses, and that they preferentially enroll in sections of courses that are known to be easy (Cohen, 1981; Feldman, 1989; Johnson, 2006).

To instructors who need favorable evaluations, such as those fighting for tenure and those hoping to renew one-year contracts, the temptation to create easier classes for the purpose of more favorable student evaluations can be difficult to resist. In fact, studies show that faculty are known for trying to influence student evaluation scores (Simpson & Siguaw, 2000). This is corroborated by evidence that adjunct faculty – temporary members of the faculty who run the risk of being replaced quickly – give higher grades to students than do more permanent faculty members (Sonner, 2000). Faculty are also more prone to providing students with higher grades than average in an environment where a "student-as-customer" viewpoint is more strongly endorsed (Stone, 1995).

Many would argue, however, that artificially increasing grades in this way is a disservice to students. Though learner-centered, active learning, the type often found in more challenging courses, can be uncomfortable to students, most agree that it is beneficial to their learning (Weimer, 2002). For example, discussion within courses is associated with better attainment of higher-order knowledge (Garside, 1996), and the amount of time spent studying outside of the classroom relates to academic achievement (McFadden & Dart, 1992). Do students, especially those who are viewed as customers, not deserve a classroom environment that provides them with a more thorough education?

For this reason, the purpose of this investigation is to examine if we can correct this classroom anomaly by providing evidence that artificial grade inflation may not necessarily impact students' evaluation of courses and instructors. Much of the literature surrounding grade inflation and student course evaluations is rooted in the late 20th century. Therefore, the current study provides important information about whether previous associations still exist today, nearly a decade or more after some of the most influential studies were published.

Previous research suggesting that easy courses produce stronger student evaluations frequently measured course ease in terms of students' grades. Some research has focused on other measures of difficulty, such as course workload, but they generally did so in terms of students' perceptions of this workload, rather than objective measures, such as the number of assignments or exams given per semester (Gillmore & Lowell, 1994; Greenwald & Gillmore, 1997; Marsh & Roache, 2000). Therefore, the current investigation adds to our understanding of the topic by not only updating the field on current trends regarding the association between grades and student evaluation scores, but also examining a wider range of variables regarding course difficulty than has previously been explored.

Method

Anonymous, end-of-semester teaching evaluations were collected from 156 students in six sections of three unique courses in the Psychology department of a large Southeastern University between 2011 and 2014. Students were asked to report on various aspects of their learning experience, including their instructor's effectiveness, their ability to think independently about course material, the level of mutual respect in the classroom, and their expected grade in the course, among other variables. Data were also collected regarding the number of exams and assignments given in each course.

Procedure

End of Semester Evaluations

Students completed end-of-semester evaluations for each class in which they were enrolled. These evaluations were innate to the course and were issued university-wide for all courses. Students received emails prompting them to complete the evaluations online. If students did not complete these evaluations before a university-chosen deadline, they received daily emails reminding them to complete the evaluations. Students also received emails from their instructor on the first and eighth days of each evaluation period. Said emails encouraged students to provide their honest feedback on the course so as to help the instructor improve the course for students in future semesters. The evaluation period each semester closed two weeks after students received the initial university email announcing the evaluation's availability. Fifty-eight percent of students, across the 6 sections, completed their evaluations.

For this evaluation, students were asked to rate their level of agreement with a number of statements on a Likert-type scale (1 = strongly disagree, 2 = disagree, 3 = somewhat disagree, 4 = somewhat agree, 5 = agree, 6 = strongly agree). They also provided open-ended feedback on the course and, finally, reported their expected grade in the course on a GPA-like scale (4 = A, 3=B, 2=C, 1=D, 0=F).

Of interest to this investigation are Likert-type variables relating to instructor effectiveness and students' experience of the course itself. As such, even though students reported on variables relating to the physical environment of the classroom, such as "How would you rate the physical environment in which you took this class, based upon your ability to see, hear, concentrate, and participate?," these variables were not included in the analyses. The full text of the variables included in the analyses can be found in Table 1 below, under the subheading "Evaluation Questions".

Course Difficulty Variables

Data were also collected regarding the number of exams, as well as the number of in-class and out-of-class assignments, administered each semester. Exams consisted of 40-50 multiple choice questions each, and each exam was administered during a single class-period, with no other class discussion happening on exam days. In-class assignments were defined as those assignments that were assigned, completed, and submitted within a single class period, with no expectation of out-of-class effort to occur on the assignment. In-class assignments could be completed in groups and were discussed at a class-wide level upon submission, during the same class period. Out-of-class assignments, or homework, were those assignments that were completed outside of the classroom and submitted on the course website associated with each class.

Results

Descriptive Statistics

Descriptive statistics for all variables of interest in this study can be found in Table 1. The reported maximum on each evaluation question score matched the maximum possible score of each scale, while the reported minimum only sometimes matched the minimum possible score of the scale. Means also fell toward the top of the scales. Numbers of in-class activities ranged between a low of 1 during one semester to a high of 21 during a different semester, while number of out-of-class homework assignments remained consistent at 2, and tests varied between 3 and 4 administrations per semester.

Pearson Correlations

Evaluation Responses and Difficulty of the Course. To test the hypothesis that the difficulty of a course would correspond to changes in student responses to instructor evaluations, Pearson correlations were calculated between evaluation questions and a student's expected grade, and the numbers of in-class activities and exams administered throughout the semester. The expected grade in the course was included as a "course difficulty" variable for the purpose of this analysis, in order to examine if students' perception of grade impacted their assessment of the course. The number of out-of-class homework assignments was excluded from this analysis, because it did not vary from semester to semester. Table 2 presents the results of the analysis.

Course difficulty variables rarely related to course evaluation responses. The expected grade in the course positively related to responses on only three evaluation questions (all other r's < .154, all other p's > .057), and the number of in-class activities assigned positively related to only a student's reported ability to think independently about course material (all other r's < .128, all other p's > .114). In contrast, the number of tests assigned in a course negatively related to five separate variables (all other r's < -.151, all other p's > .061). Expected grade, in-class assignments, and number of exams administered were all unrelated to students' agreement with the statement "Overall, the instructor's teaching was effective" (all r's < .138; all p's > .088).

Evaluation Responses and Overall Course Effectiveness. In order to determine if the above lack of correlations was indicative of a true lack of association or, rather, a lack of cohesiveness in student responses, Pearson correlations were also calculated between students' ratings on individual items on the evaluation and students' rating of overall teaching effectiveness. This teaching effectiveness variable is a single item on the evaluation (rather than a calculated average of multiple variables), but was chosen for this analysis because it is frequently used as a single number meant to represent an instructor's effectiveness at the institution at which these evaluations were given. As can be seen in Table 2, this effectiveness rating related to responses on all other evaluation questions (all other r's > .433, all other p's < .001).

Discussion

Over the last half-century, various reports have described how grades in college courses have risen despite students reporting less time spent studying (e.g., de Vise, 2012). One potential explanation for this trend is that instructors face pressure to maintain strong student evaluation scores and believe that they can achieve higher scores by decreasing the difficulty of their course and artificially inflating grades. Historical research has found that this tactic may be well founded—that there is an association between assigning high grades and earning high student evaluation scores. However, this research has limited application to today's academic climate, as some of the most important studies were conducted more than a decade ago. Moreover, very few focus on variables of course difficulty that can be objectively reported on by the instructor.

Thus, the purpose of this investigation was to determine if a course's perceived difficulty level still shows an association with student evaluation scores today. In the current investigation, each measurement of course difficulty showed some association with student responses on some evaluation items, but none related to students' ratings of overall instructor effectiveness. Meanwhile, the strength of associations between these difficulty measures was not of the same magnitude as were the associations between

individual items of the evaluation. As such, the evidence suggests that instructors need not feel pressure, as they did in the past, to reduce course demands in order to improve student evaluations.

Course Difficulty and Teaching Evaluations

One's expected grade in a course, as well as the number of assignments and exams given in each course, did impact ratings on some individual items of the course evaluation, though these correlations were relatively small and sporadic. Students' expected course grades, for example, did not relate to any items related to their view of the instructor. Instead, the expected grades related to items about themselves and their understanding of the material, such as "I have improved my ability to think independently about course material." Perhaps students are praising their instructor indirectly through these responses, but it is also possible that students who earn higher course grades actually do have a stronger ability to think independently about course material than do students who do more poorly in the class, and that the association between grades and evaluation items is warranted here and not indicative of grade inflation.

Similarly, students who completed more in-class assignments also reported a better ability to think independently about course material. If easier courses were still a strong predictor of high teaching evaluations, one would not expect this finding. Expecting more active work from students in this way can cause some grumblings (Weimer, 2002), yet some characteristic of these in-class assignments actually increased students' reports of independent thinking. One possible explanation of this could be the increased class discussion that resulted from each of these assignments, yet previous research suggests that students give lower evaluation scores in courses that involve more active

learning such as this (Lake, 2001). Perhaps, then, the positive association between number of assignments and student ratings can be explained by the feedback that accompanied these assignments, as submissions were returned to students with written comments on their work. Students appreciate feedback on assignments, indicating that they are motivated by more than just a grade, and those who receive personalized feedback in a course are more satisfied than those who do not (Gallien & Oomen-Early, 2008; Higgins, Hartley, & Skelton, 2002).

The negative association between number of exam scores and various teaching evaluation items is the only one that would have been predicted by previous research relating easier classes to higher evaluations. When students were given more exams, they reported less positive characteristics of their own learning as well as less positive characteristics of the instructor and class environment. One possible explanation for these negative associations is that exams were administered for full class periods, meaning that those courses with more exams had fewer class periods to discuss course material. More exams may also cause more test anxiety, which is known to lead to lower course performance (Hill & Wigfield, 1984; Maehr & Midgley, 1991) and perhaps lower evaluations. Perhaps the negative association lies in the nature of the exams themselves. All exams in this investigation were multiple choice with no written feedback given by instructors. Students believe that multiple choice tests measure a lower form of knowledge and adjust their study techniques accordingly (Scouller, 1998). Thus, when they are given more multiple choice exams throughout the semester, they may feel less need to deeply engage in the course and learn material, which explains the negative association between exams given and course evaluations.

With all of this being said, it is important to note that the strength of associations between course difficulty items and evaluation items is low. What's more, none of the measures of course difficulty in this investigation related to the item, "Overall, the instructor's teaching was effective", which is the item most frequently used from this array to concisely describe an instructor's ability. Therefore, this investigation reveals that in today's society, it may not be beneficial to instructors to artificially decrease the difficulty of their course in an effort to receive higher course evaluation scores.

With the pressure to artificially inflate grades removed, perhaps instructors can begin to better serve their students with more challenging courses. This transition away from the current trend will be difficult for many, as universities and courses will not have uniform levels of course difficulty, and as students in more challenging courses will have lower grades than their grade-inflated peers, thus making them less competitive for the job market (Johnson, 2006; Sabot & Wakeman-Linn, 1991). Still, instructors need to consider the benefits of creating more academically rigorous courses for their students. Over the course of the last 50 years, at the same time as students' grades steadily rose, the amount of time spent studying steadily fell, and the number of students making no gains in critical thinking throughout college rose (Arum & Roksa, 2011; de Vise, 2012). By increasing the rigor of courses, perhaps we can counteract these negative consequences of grade inflation.

Limitations and Future Directions

This investigation provides important evidence that instructors need not feel pressured to artificially decrease the difficulty of their courses in order to improve student evaluation scores. However, some limitations need to be considered. First, the correlation of individual evaluation items was quite high. Taking this in tandem with the fact that scores on most items skewed toward the top end of the distribution, there may be a halo effect in students' evaluations. Second, this investigation, relative to more prolific historical investigations, involved a relatively small sample size across a relatively homogenous sample of courses and students. More research is necessary to determine if these same effects can be found in a wider span of classes and students in which students may provide more negative course evaluations. Third, though this investigation provided evidence that a larger number of multiple choice exams in a course may relate to more negative ratings on some evaluation items, more research is necessary to determine if this association holds true across other varieties of exams (short answer, essay, mixed, etc.). Finally, because scores on evaluation items in this investigation were relatively high, future research should determine if some environments are more likely than others to discourage the historical association between high grades and high teaching evaluations. For example, perhaps students are more forgiving of rigorous academic environments when instructors demonstrate high support for students and create an environment of mutual respect.

In the meantime, the results imply that instructors under pressure of limited contracts need not artificially decrease the difficulty of their courses, but should instead work to give their students a better learning environment. Classes can be made more challenging as long as students can be provided proper feedback on assignments, and as long as exams are not given at the expense of more thorough course discussion. Perhaps in the future, instructors across disciplines and universities can find a solution to the damage that grade inflation has caused. We need to work together to help students learn effectively while remaining

competitive for graduate school and the job
market.

References

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

Bowers, J. (1970). The comparison of GPA regression equations for regularly admitted and disadvantaged freshmen at the University of Illinois. *Journal of Educational Measurement*, *7*(4), 219-225.

Cohen, P. A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Review of Educational Research*, *51*(3), 281-309.

de Vise, D. (2012, May 21). *Is college too easy? As study time falls, debate rises*. Retrieved from The Washington Post: https://www.washingtonpost.com/local/education/is-college-too-easy-as-study-time-falls-debate-rises/2012/05/21/gIQAp7uUgU_story.html

Eiszler, C. F. (2002). College students' evaluations of teaching and grade inflation. *Research in Higher Education, 43*(4), 483-501.

Feldman, K. A. (1989). Instructional effectiveness of college teachers as judged by teachers themselves, current and former students, colleagues, administrators, and external (neutral) observers. *Research in Higher Education*, *30*(2), 137-194.

Gallien, T., & Oomen-Early, J. (2008). Personalized versus collective instructor feedback in the online course room: Does type of feedback affect student satisfaction, academic performance and perceived connectedness with the instructor? *International Journal on E-learning*, *7*(3), 463-476.

Garside, C. (1996). Look who's talking: A comparison of lecture and group discussion teaching strategies in developing critical thinking skills. *Communication Education*, *45*(3), 221-227.

Gillmore, G. M., & Lowell, N. (1994). *The effects of course demands and grading leniency on student ratings of instruction.* Retrieved from the University of Washington: http://depts.washington.edu/assessmt/pdfs/reports/OEAReport9404.pdf

Greenwald, A. G., & Gillmore, G. M. (1997). No pain, no gain? The importance of measuring course workload in student ratings of instruction. *Journal of Educational Psychology*, *89*(4), 743.

Higgins, R., Hartley, P., & Skelton, A. (2002). The conscientious consumer: Reconsidering the role of assessment feedback in student learning. *Studies in Higher Education*, *27*(1), 53-64.

Johnson, V. E. (2006). *Grade inflation: A crisis in college education*. New York: Springer Science & Business Media.

Juola, A. E. (1976). *Grade Inflation in Higher Education: What Can or Should We Do?* National Council on Measurement in Education Annual Meeting, San Francisco, CA. (ERIC Document Reproduction Service No 129 917)

Lake, D. A. (2001). Student performance and perceptions of a lecture-based course compared with the same course utilizing group discussion. *Physical Therapy*, *81*(3), 896-902.

Maehr, M. L., & Midgley, C. (1991). Enhancing student motivation: A schoolwide approach. *Educational psychologist*, *26*(3-4), 399-427.

Marsh, H. W., & Roche, L. A. (2000). Effects of grading leniency and low workload on students' evaluations of teaching: Popular myth, bias, validity, or innocent bystanders? *Journal of Educational Psychology*, *92*(1), 202.

McFadden, K., & Dart, J. (1992). Time management skills of undergraduate business students. *Journal of Education for Business*, *68*(2), 84-88.

Sabot, R., & Wakeman-Linn, J. (1991). Grade inflation and course choice. *The Journal of Economic Perspectives*, *5*(1), 159-170.

Scouller, K. (1998). The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. *Higher Education*, *35*(4), 453-472.

Simpson, P. M., & Siguaw, J. A. (2000). Student evaluations of teaching: An exploratory study of the faculty response. *Journal of Marketing Education*, *22*(3), 199-213.

Sonner, B. S. (2000). A is for "adjunct": Examining grade inflation in higher education. *Journal of Education for Business*, *76*(1), 5-8.

Stone, J. E. (1995). Inflated grades, inflated enrollment, and inflated budgets: An analysis and a call for review at the state level. *Education Policy Analysis Archives*, *3*(11), 1-30.

Weimer, M. (2002). *Learner-centered teaching: Five key changes to practice*. San Francisco: Jossey-Bass.

Table 1
Descriptive Statistics for Key Study Variables

	Min	Max	Mean	SD
Evaluation Questions				
The instructor was well prepared.	3	6	5.77	.546
The instructor presented the subject matter clearly.	2	6	5.74	.645
The instructor provided feedback intended to improve my course performance.	2	6	5.66	.727
The instructor fostered an atmosphere of mutual respect.	4	6	5.84	.398
I have a deeper understanding of the subject material as a result of this course.	3	6	5.62	.733
My interest in the subject matter was stimulated by this course.	1	6	5.40	1.099
Overall, the instructor's teaching was effective.	2	6	5.62	.724
I improved my ability to think independently about the course material.	2	6	5.34	.725
I learned to identify problems and explore different solutions.	2	6	5.29	.856
The instructor used a scholarly approach in presenting content	2	6	5.58	.703
The instructor treated students with respect.	4	6	5.86	.397
The instructor was effective in administering the class and organizing materials.	2	6	5.74	.615
The grade I expect in this course is.	2	4	3.54	.550
Course Difficulty Variables				
Number of In Class Activities	1	21	7.97	8.114
Number of Homework Assignments	2	2	2.00	.000
Number of Tests	3	4	3.73	.445

Notes. N's range from 153 to 156 due to occasional missing data. Evaluation questions were rated on a 6-point Likert-type scale, though responses on some scales did not show this same range of variability. Expected course grade is student-reported and follows a typical 4.0 grading scheme (4 = A, 3 = B, …0 = F).

Table 2

Pearson Correlation of Evaluation Items with Indicators of Course Difficulty

Evaluation Questions	Effective	Grade	ICA	Exams
The instructor was well prepared.	.818***	.137	.086	-.129
The instructor presented the subject matter clearly.	.769***	.119	.128	-.179*
The instructor provided feedback intended to improve my course performance.	.697***	.154	.118	-.150
The instructor fostered an atmosphere of mutual respect.	.657***	.086	.124	-.167*
I have a deeper understanding of the subject material as a result of this course.	.840***	.180*	.081	-.136
My interest in the subject matter was stimulated by this course.	.805***	.204*	.091	-.151
Overall, the instructor's teaching was effective.	--	.133	.095	-.138
I improved my ability to think independently about the course material.	.599***	.177*	.171*	-.223**
I learned to identify problems and explore different solutions.	.577***	.159*	.123	-.177*
The instructor used a scholarly approach in presenting content	.699***	.041	.155	-.202*
The instructor treated students with respect.	.433***	.038	.027	-.059
The instructor was effective in administering the class and organizing materials.	.601***	.148	.036	-.071

Notes. N's range from 153 to 156 due to occasional missing data. "Effective" = item related to overall rating of instructor effectiveness. "Grade" = Expected course grade is student-reported and follows a typical 4.0 grading scheme (4 = A, 3 = B, …0 = F). "ICA" = number of in-class assignments. "Exams" = number of tests administered. Number of out of class homework assignments was not included in this analysis, as the number did not vary from semester to semester. * = $p < .05$, ** = $p < .01$, *** = $p < .001$.

Amanda Joyce is Assistant Professor, Murray State University.