Are Grades Inflated for Good Teaching Evaluations?
Evidence From Bangladesh[*]

Gour Gobinda Goswami, Abdul Mumit
North South University, Dhaka, Bangladesh

The hazard of excessive reliance on using student evaluations of teaching effectiveness, namely, teaching evaluations, has been found to be a contributing factor towards grade inflation. As performance evaluation of faculty members are largely dependent on these teaching evaluations, there is a significant incentive for course instructors to give higher, potentially inflated, grades to students in return for favourable teaching evaluations as positive reciprocity. The likelihood of grade inflation is further exacerbated for part-time and adjunct faculty members, who rely more on good teaching evaluations than their full-time compatriots. This study investigated the question of whether the use of teaching evaluations has been a contributing factor towards grade inflation at a certain private university in Bangladesh. Analysing the data for 98,362 student evaluations of more than 3,653 course sections, taught by more than 350 faculty members, this study examined the relationship between teaching evaluation and average class grade for courses offered during Spring 2012, Summer 2012, and Fall 2012 semesters at this university. As a novel econometric analysis in such research, instrumental variable estimation is undertaken, whereby statistically significant results indicate a positive relationship between teaching evaluation and average class grade for courses, after controlling for latent variables, such as teaching quality and other independent variables. The conclusion that higher grades beget high evaluation scores regardless of teaching quality indicates grade inflation for ensuring good teaching evaluation.

Keywords: grade inflation, teaching evaluation, teaching quality

Introduction

Grade inflation, across multiple levels of education, continues to receive active attention from both the scholarly community and general press. Grade inflation leads to compression of grades in the higher categories and thus diminishes their value as an indicator of student abilities. The hazard of grade inflation is that without it, a truly exceptional student might be awarded an A, while a very good student might receive a B+, thus, ensuring discriminating validity. However, with grade inflation, both students receive as, making it hard for employers and graduate schools to differentiate them. Grave negative consequences of grade inflation in higher education has been extensively studied from various perspectives, encompassing how it disincentivizes students to give maximum effort (Sabot & Wakeman-Linn, 1991; Pattison, Grodsky, & Muller, 2013), inclination of

students to take leniently graded courses only (Bar, Kadiyali, & Zussman, 2009), and imposition of social efficiency costs (Herron & Markovich, 2017). However, in the context of Bangladesh with a nascent higher education sector, grade inflation engenders additional repercussions. Good students from Bangladesh universities commonly set their goals on pursuing higher education from foreign educational institutions, and the best students rationally hope for admission in the best institutions. However, universally high grades entail that foreign universities have difficulty selecting the best students and makes separation of superior performers from the everyday, average student more difficult. Overall, this takes away credibility from the grading system and demotivates students from maximizing learning efforts.

Inquiry into the causes for grade inflation is extant (see L. W. Lackey & W. J. Lackey, 2006, for a summary). One particular posited cause, commonly stated as the "grade leniency" hypothesis, is that faculty members influenced by desire for good teaching (course) evaluations by students, offer inflated grades (Wilson, 1998; Rosovsky & Hartley, 2002). In a summative study, Marsh and Roche (2000) reported the average correlation between expected grades and teaching evaluations + 0.20. However, correlational studies have indicated positive relationship between teaching evaluations and average class grade, causal inference have not been made (Eiszler, 2002). The major hindrance towards causal inference is the possible existence of the latent variable "teaching quality", as good teaching would result in good grades and good evaluations. In statistical terms, this gives rise to an "endogeneity problem", whereby the underlying latent variable (teaching quality) affects both the dependent (teaching evaluation) and independent variable (course grades). This research implements a novel approach by undertaking an instrumental variable estimation method, to control for the endogenous effect of teaching quality, and thus, attempts to establish a causal link between teaching evaluation and course grades.

Aggregate data for 98,362 teaching evaluations done by students, enrolled in 3,653 sections for courses, taught by more than 350 faculty members, offered during Spring 2012, Summer 2012, and Fall 2012 semesters (sometimes also denoted as trimesters in North American universities) is analysed for the leading private university in Bangladesh, to which anonymity has been granted intentionally to minimize confirmation and other sources of bias. The size of the dataset designates a special significance to this study in terms of generalization to various other tertiary educational institutions. Despite statistical controls to eliminate endogeneity bias, statistically significant results indicate a positive relationship between teaching evaluation and average class grade. This predictive relationship between student ratings of teaching and average class grades is significant even after variables related to alternative explanations have been statistically controlled. The rest of the research is structured as follows. In Section 2, we attempt to identify factors that influence teaching evaluation (in addition to course grades), Section 3 explicates the methodology, Section 4 describes the data, Section 5 reports the results of the quantitative analysis, Section 6 discusses the results, Section 7 prescribes policy recommendations, Section 8 supplements the numerical results with qualitative analysis, and Section 9 concludes.

Determinants of Teaching Evaluation

This research utilizes the classification scheme of Davies, Hirschberg, Lye, Johnston, and McDonald (2007), by identifying determinants of teaching evaluations scores across the following three broad categories: faculty-level determinants, student-level determinants, and subject-level determinants.

Faculty-Level Determinants

Course evaluations exhibit a negative relationship with increasing age and job duration in the form of number of years of teaching experience (Wolfer & Johnson, 2003). However, title, degree, and job position of the faculty (teaching assistant, visiting assistant, and associate or full-rank tenure-track professor) statistically do not affect teaching evaluations of them. The reputation of the faculty, in terms of his/her reputation for teaching quality, significantly influences teaching evaluations, whereby instructors that enjoy good reputations also receive relatively high evaluations (Griffin, 2001). Gender of the faculty generally does not bias teaching evaluations, although in a minority of studies male teachers receives higher overall evaluations than did female instructors (Feldman, 1992).

Student-Level Determinants

Gender of the student did not have statistically significant consequence on teaching evaluations. Centra (1993) found that the age of a student does not predispose them in terms of evaluating instructors. In contrast, Worthington (2002) reported that the age of the student did have some influence on the ratings given. Students, who are over the age of 30 and were also female, are more likely to assign a lower rating to the instructor. Teaching evaluations are positively affected by the level of academic aptitude of students (Mason, Steagall, & Fabritius, 1995).

Subject-Level Determinants

Marsh (1987) found diminutive evidence of the influence of class size on teaching evaluation of that class. In a more recent study, Neumann (2000) noted that larger classes give lower teaching evaluations. Evaluations of teaching quality are found to be unrelated to the part of the year when the course is conducted (Liaw & Goh, 2003). The academic discipline of the course, in terms of humanities, arts, science, etc. affects the teaching evaluations. Feldman (1978) found humanities and arts subjects receive higher ratings than social science courses. The lowest ratings are reserved for mathematics courses. Furthermore, the level of the course also positively correlates with teaching evaluations. Third and fourth year courses tend to receive higher ratings in relation to first-year courses (Marsh & Bailey, 1993).

Apart from these determinants, grade leniency has also been identified as a possible determinant of teaching evaluation scores. The leniency hypothesis indicates that students give higher teaching evaluation scores to instructors from whom they receive good grades. From this point of view, teachers can "buy" better evaluations by grading in a lenient manner. Students thus reward their instructors for assigning good course grades, and conversely retaliate with lower teaching evaluations for low course grades. Krautmann and Sander (1999) showed that a one-point increase in the expected classroom grade point average (GPA) leads to an improvement of between + 0.34 and + 0.56 in the teaching evaluation score. McPherson and Jewell (2007) showed that an increase of one point (on a 4-point scale) in the student's expected grade leads to an increase of approximately + 0.25 (on a 4-point scale) in the teaching evaluation score. A more recent study (McPherson, Jewell, & Kim, 2009) confirmed that instructors can ensure higher evaluation scores by inflating students' grade expectations.

Therefore, along with the instructor, student, and subject level determinants, we also posit average class grade as a determinant of teaching evaluation scores for our subsequent analysis. In the following sections, the key findings are reported and the implications for future practice are explored.

Methodology

Instrumental variable estimation is employed, which is a novel econometric technique in grade inflation research. For multivariate regression, generalized linear modelling (GLM) is used, which area significant class of statistical models for relating responses to linear combinations of predictor variables, including many commonly encountered types of dependent variables and error structures as special cases. Specifically, when the asymptotic assumptions of the standard general linear model prove to be extraneous due to the nature of the data, GLM offers flexibility in terms of distributional structure of the response variable, non-linear modelling of the relationship between the response and the independent variables etc. We closely follow the canonical treatment of GLMs as Nelder and Wedderburn (1972). Firstly, we note that the response variable (teaching evaluation scores) exhibits two distinct characteristics: The variable is continuous in nature (more than 30 categories) and only takes strictly positive values due to its definitional method. In Unison, these characteristics indicate towards a "gamma" distribution of the response variable. Therefore, we estimate our model assuming a gamma distributed response variable and the canonical link function. We incrementally introduce the various categories of independent variables and sequentially evaluate the change in magnitude and significance of coefficient estimates.

Control for Endogeneity and Instrumental Variable Approach

A major complication that is emphasized in micro-econometrics is the possibility of inconsistent parameter estimation due to endogenous regressors, which can give rise to inconsistent and biased parameter estimates. For instance, in our analysis, the latent variable of "teaching quality" can cause both high grades (as students are motivated to work hard) and high teaching evaluation scores (students reward the course instructor aptly for the high teaching quality). Presence of such a latent variable would lead to spurious correlation between teaching evaluation scores and class grades, leading us erroneously to conclude evidence of grade inflation.

The most commonly used econometric method to model data that are affected by the unobservable confounding issue is the instrumental variable (IV) approach (Wooldridge, 2002). This method can yield consistent parameter estimates and can be used in any kind of analysis in which unmeasured confounding is suspected to be present (Linden & Adams, 2006). In the research, we identify the number of sections taught by a course instructor in a particular semester as an instrument for teaching quality. It is rationalized that course instructors with relatively good teaching quality would be allowed to teach more sections, as it is sub-optimal for the administration to maximize students' exposure towards course instructors of relatively lower teaching quality. Furthermore, ensuing qualitative discussions also confirm the instrument of number of courses taught as an accurate indicator of teaching quality, whereby good teaching is rewarded by approval to teach more sections. Conversely, relatively poor teaching quality necessitates departmental heads to curtail the number of sections taught. Furthermore, the number of sections taught is not expected to be correlated with the other regressors. The qualitative discussions following the methodology of (Marra & Radice, 2010), we implement a modified 2 SLS (2 stage least squares) regression method to estimate relevant parameters. Firstly, we regress the number of courses taught in a semester on the average course grades, in a GLM structure, assuming a Gaussian response variable and the canonical link function. In the next step, the expected values of class grades are used as a regressor in the complete model with all the other independent variables.

Data and Variable Description

Feistauer and Richter (2017) suggested that the reliability of teaching evaluations is ensured when aggregated evaluations based on a sufficient number of students are used. In that connection, the dependent variable is the teaching evaluation score, which is calculated as the average of teaching evaluation scores received across Question # 1-14 in the teaching evaluation questionnaire. Likert scale responses have been converted to numeric values by equating 5 to "Strongly agree" response, 4 to "Agree" response, 3 to "Neutral" response, 2 to "Disagree" response, and 1 to "Strongly disagree" response.

The independent variables are grouped into three major categories: section specific variables, instructor specific variables, and course specific variables. In evaluating the relationship between teaching evaluation and grades, we should note that final grades in a course are typically not known to students at the time they complete a student evaluation form, and thus, should not be expected to bias their evaluations. Forms are generally filled out during the last few weeks of a course when students are aware only of the grades they think they will get based on their perceived performance to date. Therefore, studies of a possible grading-leniency bias on ratings should appropriately use expected grades and not actual grades (as we have done). However, an extensive body of research, reviewed by (Franklin & Theall, 1991), indicate that students' expected grades tend to show the same associations with other variables as do actual final grades. Therefore, we use actual grades as a proxy for expected grades in this research.

Table 1 provides an overview of the variables and related summary statistics.

Table 1

Variable Description and Summary

Variable	Description	Arithmetic Mean (μ)	Standard deviation (σ)
Teaching evaluation	Average of teaching evaluation scores received across Question # 1-14 in the teaching evaluation questionnaire. Likert scale responses were converted to numeric values by equating 5 to "Strongly agree" response and 1 to "Strongly disagree" response, with intermediary numeric values assigned to intermediate responses.	4.18	0.50
Section specific variables			
Average class grade	Average of letter grades received by students in a particular section. Letter grades were converted to grade points as per the university's grading policy.	2.59	0.62
Proportion of students with grades A- and above	The number of students in a particular section who received a letter grade equal to or higher than A- ÷ Total number of students in that section	0.24	0.18
Proportion of students with grades D and below	The number of students in a particular section who received a letter grade equal to or lower than D ÷ Total number of students in that section	0.16	0.16
Class size	The number of students in a particular section	26.93	11.81
Instructor specific variables			
Job duration at the university	Tenure of job (in months) till October 2012 since date of joining	18.59	12.86
Gender	Dummy variable = 1, for sections taught by male faculty members. Reference category is female.	N/A	N/A
Full-time faculty member	Dummy variable = 1, for sections taught by full time faculty members. Reference category is part-time faculty members.	N/A	N/A
Visiting faculty member	Dummy variable = 1, for sections taught by visiting faculty members. Reference category is part-time faculty members.	N/A	N/A

(Table 1 to be continued)

Junior lecturer	Dummy variable = 1, for sections taught by Jr. lecturers. Reference category is professors.	N/A	N/A
Lecturer	Dummy variable = 1, for sections taught by lecturers. Reference category is professors.	N/A	N/A
Senior lecturer	Dummy variable = 1, for sections taught by Sr. lecturers. Reference category is professors.	N/A	N/A
Assistant professor	Dummy variable = 1, for sections taught by assistant professors. Reference category is professors.	N/A	N/A
Associate professor	Dummy variable = 1, for sections taught by associate professors. Reference category is professors.	N/A	N/A
The number of sections taught per semester	The number of sections taught by a faculty member in a specific semester.	4.87	3.07
Course specific variables			
200 Level Course	Dummy variable = 1, for sections of 200 level course. Reference category is sections of 100 level course.	N/A	N/A
300 Level Course	Dummy variable = 1, for sections of 300 level course. Reference category is sections of 100 level course.	N/A	N/A
400 Level Course	Dummy variable = 1, for sections of 400 level course. Reference category is sections of 100 level course.	N/A	N/A
500 Level Course	Dummy variable = 1, for sections of 500 level course. Reference category is sections of 100 level course.	N/A	N/A
600 Level Course	Dummy variable = 1, for sections of 600 level course. Reference category is sections of 100 level course.	N/A	N/A
700 Level Course	Dummy variable = 1, for sections of 700 level course. Reference category is sections of 100 level course.	N/A	N/A
800 Level Course	Dummy variable = 1, for sections of 800 level course. Reference category is sections of 100 level course.	N/A	N/A
School of Engineering	Dummy variable = 1, for sections taught by the School of Engineering. Reference category is sections taught by School of Business & Economics.	N/A	N/A
School of Life Sciences	Dummy variable = 1, for sections taught by the School of Life Sciences. Reference category is sections taught by School of Business & Economics.	N/A	N/A
School of Humanities	Dummy variable = 1, for sections taught by the School of Humanities. Reference category is sections taught by School of Business & Economics.	N/A	N/A
Summer 2012	Dummy variable = 1, for sections pertaining to Summer 2012 semester. Reference category is Spring 2012 semester.	N/A	N/A
Fall 2012	Dummy variable = 1, for sections pertaining to Fall 2012 semester. Reference category is Spring 2012 semester.	N/A	N/A

Firstly, we note that the arithmetic mean of teaching evaluation scores is fairly high (4.18), which entails that students on average respond between "Strongly agree" and "Agree" to various questions regarding the quality of the teaching during the semester and individual characteristics of the faculty member. The fairly low standard deviation (0.50) further indicates that the predilection towards providing positive teaching evaluation scores is fairly consistent across various dimensions of academic disciplines, semesters, etc.

The arithmetic mean of average class grade is only 2.59, which is just below a B- (minus) letter grade. The standard deviation is relatively large at 0.62. This indicates that the average class grade varies considerably across various sections of academic discipline. On average 24% of students receive a grade of A- (minus) or above, and 16% students receive a grade of D or below, across all sections. The former result especially indicates towards a tendency of "easy grading" whereby a relatively large proportion of the section receives the top grades.

The average job duration of faculty members is 18.5 months (around 1.5 years) which exhibits the young complexion of the faculty body, whereby most teaching is done by lecturers. In fact, nearly 54.01% of all sections are taught by senior lecturers, lecturers and junior lecturers. 18 % courses are taught by full professors.

To delve further into the data for gathering additional insights, we cross tabulate teaching evaluation scores across certain dimensions. Table 2 summarizes these results.

Table 2

Cross Tabulation Results for Teaching Evaluation Scores

Variables		Arithmetic mean of teaching evaluation scores	Standard deviation of teaching evaluation scores
Degree of engagement of faculty member	Full-time faculty member	4.02	0.50
	Visiting faculty member	4.18	0.51
	Part-time faculty member	4.20	0.46
Academic discipline	School of Business	4.18	0.49
	School of Engineering	4.13	0.48
	School of Life Sciences	4.24	0.52
	School of Humanities	4.17	0.50
Ranking of faculty member	Junior lecturer	4.07	0.60
	Lecturer	4.16	0.47
	Senior lecturer	4.24	0.47
	Assistant professor	4.14	0.59
	Associate professor	4.17	0.53
	Professor	4.27	0.42

In analysing the results across various categories of faculty members (full-time/part-time/visiting), we notice that full-time faculty members receive 4.5% lower teaching evaluation scores than part-time faculty members, and 4.0% lower teaching evaluation scores than visiting faculty members. If teaching evaluation is indeed seen as an unbiased and consistent estimator teaching quality, then the results suggest that part-time faculty members are better pedagogues in comparison to their visiting and full-time counterparts. However, this interpretation can be negated by the fact that part-time faculty members usually teach relatively easy, lower level courses, which allow them to secure relatively higher teaching evaluation scores. It could also be symptomatic of grade inflation, whereby higher grades are offered to receive better evaluation scores.

There is very little systematic variation of teaching evaluation scores across various academic disciplines. It is indicative towards absence of any systematic bias in teaching evaluation scores across the various schools of the university. In analysing the variation of teaching evaluation scores across ranking of faculty members, we note that senior lecturers and professors exhibit distinct superiority in terms of teaching evaluation scores. Again, if teaching evaluation scores are accepted as a fair estimator of teaching quality, then higher than average scores of professors and senior lecturers bode well as it exhibits commitment to teaching quality from the senior most ranks of pre-Ph.D. (senior lecturers) and post-Ph.D. (professors) faculty members. With the alternative explanation of grade inflation, it could also be seen under negative light, whereby these senior-most ranks deliberately inflate grades to assure higher than average teaching evaluation scores. Furthermore, we apportion our interest towards variation of average class grade across various dimensions. Table 3 summarizes these results.

In analysing the results across various categories of faculty members (full-time/part-time/visiting), we notice that in sections taught by visiting faculty, students receive lower grades on average. Average grades received by students in sections taught by part-time and full-time faculty members are nearly identical in mean and variance. This result partially allays fears of "easy grading" being undertaken by part-time faculty members.

Table 3

Cross Tabulation Results for Average Class Grade

Variables		Arithmetic mean of average class grade	Standard deviation of average class grade
Degree of engagement of faculty member	Full-time faculty member	2.58	0.62
	Visiting faculty member	2.38	0.71
	Part-time faculty member	2.60	0.61
Academic discipline	School of Business	2.58	0.61
	School of Engineering	2.47	0.61
	School of Life Sciences	2.65	0.62
	School of Humanities	2.68	0.62
Ranking of faculty member	Junior lecture	2.51	0.66
	Lecturer	2.45	0.56
	Senior lecturer	2.56	0.62
	Assistant professor	2.65	0.66
	Associate professor	2.75	0.61
	Professor	2.74	0.62

Comparing the results across the different academic disciplines, engineering students appear to receive average class grades lower than that of their cohorts in other schools. This could be symptomatic of engineering being a relatively more difficult subject than the other academic disciplines, or it could also indicate towards lower quality of studentship in the engineering department.

In analysing the variation of average class grades across ranking of faculty members, we note that faculty members in the professorial rank (assistant, associate, or full) award higher average class grades than their colleagues of lower ranks. Apparently, lecturers are the thriftiest in terms of offering higher grades.

Results

Both quantitative and qualitative methods were used to gain insights on the data. In the ensuing sections, we focus on the quantitative and qualitative findings respectively.

Table 4 presents the results of the 2 SLS estimation. Only variables which have a statistical significance level of $p < 0.10$ are reported for the sake of brevity.

Firstly, we notice that out of three groups of independent variables, section specific variables have the strongest influence on teaching evaluation scores. Most of the other independent variables from the other two categories do not have a statistically significant impact on teaching evaluation scores.

We primarily observe that a statistically significant positive relationship between average class grades and teaching evaluation scores is observable. This relationship is quite robust, as the magnitude of the impact remains fairly stable across various specifications. The proportion of students with grades A- and above have a significantly positive impact on teaching evaluation and concurrently proportion of students with grades D and

below has a significantly negative impact. These findings are also remarkably robust. Larger class sizes also positively affect teaching evaluation scores.

Table 4

Estimation Results for Aggregated Teaching Evaluation Scores

Independent variables		Specification (1)	Specification (2)	Specification (3)
Section specific variables	Average class grade	0.13^{*}	0.230^{***}	0.167^{**}
	Proportion of students with grades A- and above	0.433^{***}	0.445^{***}	0.445^{***}
	Proportion of students with grades D and below	-0.284^{***}	-0.260^{***}	-0.258^{***}
	Class size	0.005^{***}	0.004^{***}	0.004^{***}
Instructor specific variables	Job duration at the university		0.003^{***}	0.004^{***}
	Full-time faculty member		-0.057^{***}	-0.054^{***}
Course specific variables	School of Engineering			-0.055^{***}
	Summer 2012			0.038^{*}
	AIC	4.864	4.869	4.875
	BIC	-29868	-29795	-29696

Notes. * denotes significance at $\propto = 0.10$, ** denotes significance at $\propto = 0.05$, *** denotes significance at $\propto = 0.01$.

In terms of instructor specific variables, course instructors with longer experience tend to receive higher evaluation scores. Interestingly, compared to part-time faculty members, full-time faculty members receive lower teaching evaluation scores. With respect to course specific variables, School of Engineering apparently receives relatively low teaching evaluation scores in comparison to the School of Business and Economics. Furthermore, teaching evaluations undertaken in Summer 2012 semester tend to be slightly higher than Spring 2012 semester.

The relationship between average class grade and teaching evaluation scores is significantly positive and robust across various specifications. Therefore, an increase in average class grade is correlated with an increase in teaching evaluation score. To evaluate the numerical magnitude of this relationship, we compute the average marginal effect of average class grade on teaching evaluation score. Following Greene (2003), we calculate the average marginal effect as:

$$Average\ Marginal\ Effect\ (AME)$$
$$= g\ \{X_i(\bar{x}_i), X_{-i}(\bar{x}_{-i} + 1)\} - g\ \{X_i(\bar{x}_i), X_{-i}(\bar{x}_{-i})\}$$
$$where\ X_{-i} =\ Average\ class\ grade, X_i$$
$$=\ Other\ independent\ variables, g\ (.)\ is\ the\ canonical\ link\ function$$

Specifically, the dependent variable (teaching evaluation score) is evaluated at the average value of all the independent variables. Consequently, the value of a single independent variable (average class grade) is incremented by one (from its average value) while keeping the other independent variables fixed at their respective average values. The dependent variable is then evaluated at this incremented value of the independent variable and its difference with the former evaluated value is the average marginal effect (AME). The AME identifies the marginal impact of a 1.0 increase in average class grade on teaching evaluation score. Table 5 reports the AME of the independent variables that are significant at 99% level of confidence:

Table 5

Average Marginal Effect (AME) of independent variables on teaching evaluation scores

Independent variables	Average marginal effect
Average class grade	0.17
Proportion of students with grades A- and above	0.45
Proportion of students with grades D and below	- 0.25
Class size	0.005
Job duration at the university	0.004
Full-time faculty member	- 0.06
School of Engineering	- 0.05
Summer 2012	0.04

In analysing the AME, we notice that "proportion of students with grades A- and above" has the relatively largest magnitude of positive impact on teaching evaluation score, whereas "Job duration at the University" has the relatively smallest magnitude of positive impact on teaching evaluation score. Furthermore, "proportion of students with grades D and below" has the relatively largest magnitude of negative impact on teaching evaluation score, whereas being a "full-time faculty member" has the relatively smallest magnitude of negative impact on teaching evaluation score. Specifically, an increment of 1.0 in average class grade leads to an increase of + 0.17 points in teaching evaluation score.

Discussion

The positive correlation between teaching evaluation and grades is generally explained in the literature by the presence of a latent variable, which influences both teaching evaluation and grades. The primary explanation is posited through teaching effectiveness, whereby teaching effectiveness influences both grades and ratings. Good instructors teach courses in which students learn much due to which they earn deservedly high grades and correspondingly give appropriately high ratings to the course and to the instructor. By implementing instrumental variable estimation, we have eliminated the endogenous effect of teaching quality. Therefore, the positive relationship between teaching evaluation scores and average class grade that we find evidence of is purely indicative of grade inflation.

Alternatively, student motivation has been suggested as a latent variable, whereby courses that attract highly motivated students should have higher average grades (because the students work harder) and should get higher ratings (because the motivated students appreciate both course and instructor). Our research discredits this particular line of reasoning as we find no statistically significant variation in teaching evaluation scores across various course levels.

Social psychology theory predicts that students with high grades shall attribute it to diligence or intelligence whereas for low grades students can reduce cognitive dissonance and engage in ego defence by giving low evaluations to teachers who give them low grades (Maurer, 2006). Our research finds evidence for this as we note sections with relatively higher proportion of weak students (with grades of D and below) have relatively lower teaching evaluation scores.

We also note that a larger class size generally leads to a higher teaching evaluation score. This relates to the findings of Monks and Schmidt (2010), whereby large classes seem to instigate faculty members to alter their courses in ways deleterious to students and consequently lead to grade inflation through grade leniency.

The job duration of the faculty member also correlates positively with teaching evaluation scores. From an organization perspective, this bodes well, as on-the-job learning and experience of the faculty member seems to gradually improve the teaching quality.

Policy Implications

From the perspective of the university, presence of grade inflation indicates that increased institutional emphasis on teaching evaluations for job performance evaluation of faculty members can exacerbate the problems of grade inflation and inadvertently lower faculty teaching effort. This warrants a more comprehensive and inclusive set of evaluation criteria that encompass not only the teaching performance of the faculty members, but also their performance in research undertakings, commitment to university, etc. Especially, as research output is relatively easier to quantify and standardize, it is foreseen that increased emphasis on research productivity shall provide strong incentives against grade inflation.

The propensity towards grade leniency should lead to strong reputation effects, whereby faculty members that tend to grade leniently shall be the subject of interest of students who wish to receive a lenient grade and offer a good teaching evaluation score in reciprocal. Therefore, these students would tend to flock towards faculty members who have a reputation for grade inflation. This creates a virtuous cycle, as grouping together of these students will only help fuel the propensity of the faculty member to continue to inflate grades. To break this cycle, the name of the faculty members could be made hidden during the course advising process, such that students who are looking for a lenient grade cannot deliberately exert their choice of course instructor. If an instructor finds a random assortment of students, then instead of undertaking lenient grading for receiving good teaching evaluation scores, the optimal strategy for the instructor would be to maximize the quality of teaching which should translate to a good teaching evaluation. At end, there would be less incentive to practice grade inflation for faculty members who are prone to such.

Lack of systematic bias across various departments and schools of the university indicate the non-necessity of department specific policy prescriptions to combat grade inflation. This ensures that a conjoint policy to diminish the tendency for grade inflation can be implemented throughout the university, and need not be customized for each department.

The positive correlation between job duration and teaching evaluation also allows us to prescribe that the University should expend effort on retaining faculty members with relatively longer job experience. Specifically, senior lecturers tend to enjoy significantly higher teaching evaluation scores than other faculty ranks. This stands as a testament to the effort expended by senior lecturers in providing a positive learning experience for the students, and hence warrants special attention from the policy-makers.

Qualitative Analysis

An open discussion session was undertaken to delve further into the quantitative findings and also to identify future avenues of investigation. The discussants were respective Chairmen (departmental heads) of various departments at a private university in Bangladesh, representative of the various disciplines at a university, namely, Business and Economics, Humanities and Social Sciences, Engineering and Physical Sciences, and Health and Life Sciences. The position of a Chairman offered the discussants a unique and direct insight into the grading policy and procedures at their respective departments. The discussion provided the following collective insights.

Faculty members who do not receive good evaluation scores are not necessarily bad teachers. Notably, students who receive relatively low grades in the courses of these teachers with relatively high level of teaching expertise, simply attribute their grades to the teaching quality. This insight is significantly aligned to the self-attribution bias explanation for grade inflation, whereby students attribute successful outcomes to their own skill but blame unsuccessful outcomes on external factors.

It has also been observed that the faculty member's behavior, communication skill, and cooperativeness with the students' do have a significant impact in the teaching evaluation scores. Since the available data for this research does not encompass these individual characteristics of faculty members, a future avenue for research could be to measure these individual characteristics through primary surveys to determine their exact explanatory nature in the grade inflation phenomenon.

The specific policy prescription of not revealing the instructor for any particular course has proved to be ineffective from past experiences. Upon commencement of lectures in a semester, the students have a strong tendency to switch to their preferred faculty member's course, thus creating a substantive administrative burden and consequently negate the desired effectiveness of the policy.

The policy prescription of implementing a grade distribution policy uniformly across the various departments of the university was deemed to be an effective tool to abate incidences of grade inflation.

In summary, the discussion session offered some new avenues of future research (including individual characteristics of faculty members in the analysis), while largely corroborating the quantitative findings and policy prescriptions offered by the researchers.

Conclusions

This paper implements instrumental variable estimation, as a novel econometric technique in grade inflation research, to find robust and significant evidence for grade inflation while taking into consideration various course, instructor and subject related variables. It is noted that increased institutional emphasis on teaching evaluations can exacerbate the problems of grade inflation and inadvertently lower faculty teaching effort.

However, in advocating policies against grade inflation, adverse consequences of this unilateral disarmament must also be surmised. Systematically, lower grades may disadvantage them relative to students at other institutions when they present their grades to those outside the college. For instance, most jobs in Bangladesh require a minimum threshold GPA, and even use GPA as an indicator of the applicant's merit. In these cases, the university that discourages grade inflation unilaterally shall put its graduates at peril in the marketplace. On the other hand, a sustained policy of restricting grade inflation shall also benefit the institution in the long run, as the reliability of its students' grades shall be firmly established after a certain time period. In the short run, as grade inflation is a systemic problem leading to inefficient allocation of resources, universities should consider acting together in response.

From a methodological perspective, the causal link between grade leniency and teaching evaluation can be explored further through an alternative measure of the grade leniency variable. Grade leniency for a specific section can be derived as the difference between average GPA of that course (based on the offered grades) and the actual average GPA of all the students in that section. The current depth of data of this research does not allow us to formulate grade leniency in such fashion. However, future research with such a direct measure of grade leniency would allow more rigorous and robust econometric analysis.

Although the current research answers important questions regarding the relationship between grade inflation and teaching evaluation, it also raises multitudinous subtle inquiries into the finer aspects of the topic. For instance, whether this positive relationship found in this study persists through diverse cross-sections and time-periods remains to be seen, and could be a subject of further investigation. Furthermore, a secondary and objective measure of teaching quality can be used to correlate actual teaching evaluation results with this measure of teaching quality. Such a measure could be based on peer-based measures of faculty members' teaching aptitude. A lack of correlation between these two, and a positive correlation of teaching evaluation scores with expected class grades, would provide significantly robust proof of grade inflation. Lastly, the current study postulates certain policy prescriptions for negating grade inflation. Effective implementation of such policies, with ensuing evaluation of their impact would also provide rich avenues of future research.

References

Bar, T., Kadiyali, V., & Zussman, A. (2009). Grade information and grade Inflation: The Cornell experiment. *Journal of Economic Perspectives, 23*(3), 93-108.

Centra, J. A. (1993). *Reflective faculty evaluation: Enhancing teaching and determining faculty effectiveness*. San Francisco: Jossey-Bass.

Davies, M., Hirschberg, J., Lye, J., Johnston, C., & McDonald, I. (2007). Systematic influences on teaching evaluations: The case for caution. *Australian Economic Papers, 46*(1), 18-38.

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

Feistauer, D., & Richter, T. (2017). How reliable are students' evaluations of teaching quality? A variance components approach. *Assessment & Evaluation in Higher Education, 42*(8), 1263-1279.

Feldman, K. A. (1978). Course characteristics and college students' ratings of their teachers: What we know and what we don't. *Research in Higher Education, 9*, 199-242.

Feldman, K. A. (1992). College students' views of male and female teachers: Part I Evidence from the social laboratory and experiments. *Research in Higher Education, 33*, 317-351.

Franklin, J. L., & Theall, M. (1991). Grade inflation and student ratings: A closer look. Paper presented at *The 72nd annual meeting of the American Educational Research Association*, ERIC # ED 349 318.

Greene, W. H. (2003). *Econometric analysis* (5th ed.). Prentice Hall: Upper Saddle River.

Griffin, B. W. (2001). Instructor reputation and student ratings of instruction. *Contemporary Educational Psychology, 26*, 534-552.

Herron, M. C., & Markovich, Z. D. (2017). Student sorting and implications for grade inflation. *Rationality and Society, 29*(3), 355-386.

Krautmann, A. C., & Sander, W. (1999). Grades and student evaluations of teachers. *Economics of Education Review, 18*(1), 59-63.

Lackey, L. W., & Lackey, W. J. (2006). Grade inflation: Potential causes and solutions. *International Journal of Engineering Education, 22*(1), 130-139.

Liaw, S. H., & Goh, K. L. (2003). Evidence and control of biases in student evaluations of teaching. *International Journal of Educational Management, 17*(1), 37-43.

Linden, A., & Adams, J. L. (2006). Evaluating disease management programme effectiveness: An introduction to instrumental variables. *Journal of Evaluation in Clinical Practice, 12*, 148-154.

Marra, G., & Radice, R. (2010). Penalised regression Splines: Theory and application to medical research. *Statistical Methods in Medical Research, 19*, 107-125.

Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research, 11*, 253-388.

Marsh, H. W., & Bailey, M. (1993). Multidimensional students' evaluations of teaching effectiveness. *Journal of Higher Education, 64*(1), 1-18.

Marsh, H., & Roche, L. (2000). Effects of grading leniency and low work-load on students' evaluations of teaching: Popular myth, bias, validity, or innocent bystanders? *Journal of Educational Psychology, 92*(1), 202-227.

Mason, P. M., Steagall, J. W., & Fabritius, M. M. (1995). Student evaluations of faculty: A new procedure for using aggregate measures of performance. *Economics of Education Review, 14*, 403-416.

Maurer, T. W. (2006). Cognitive dissonance or revenge? Student grades and course evaluations. *Teaching of Psychology, 33*(3), 176-179.

McPherson, M. A., & Jewell, R. T. (2007). Levelling the playing field: should student evaluation scores be adjusted? *Social Science Quarterly, 88*, 868-881.

McPherson, M. A., Jewell, R. T., & Kim, M. (2009). What determines student evaluation scores? A random effects analysis of undergraduate economics classes. *Eastern Economic Journal, 35*, 37-51.

Monks, J., & Schmidt, R. (2010). *The impact of class size on outcomes in higher education.* Cornell Higher Education Research Institute, Cornell University, School of Industrial and Labor Relations.

Nelder, J., & Wedderburn, R. (1972). Generalized linear models. *Journal of the Royal Statistical Society, 135*(3), 370-384.

Neumann, R. (2000). Communicating student evaluation of teaching results: Rating interpretation guides. *Assessment & Evaluation in Higher Education, 25*(2), 121-134.

Pattison E., Grodsky E., & Muller C. (2013). Is the sky falling? Grade inflation and the signalling power of grades. *Educational Researcher, 42*, 259-265.

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

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

Wilson, R. (1998). New research casts doubt on value of student evaluations of professors. *Chronicle of Higher Education, 44*, 12-14.

Wolfer, T. A., & Johnson, M. M. (2003). Re-evaluating student evaluation of teaching: The teaching evaluation form. *Journal of Social Work Education, 39*(1), 111-121.

Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. Cambridge: MIT Press.

Worthington, A. C. (2002). The impact of student perceptions and characteristics on teaching evaluations: A case study in finance education. *Assessment and Evaluation in Higher Education, 27*(1), 49-64.