



**VOL. 20, Nº 2 (Mayo-Agosto 2016)**

ISSN 1138-414X (edición papel)

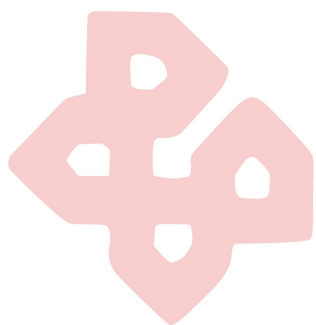
ISSN 1989-639X (edición electrónica)

Fecha de recepción 21/01/2015

Fecha de aceptación 25/09/2015

## **¿ES LA EVALUACIÓN DEL PROFESOR POR LOS ESTUDIANTES UN INCENTIVO PERVERSO?**

*Is student evaluation of teacher a wrong incentive?*



**Yuri Gorbaneff**

MSc, Professor, Management Department  
Pontificia Universidad Javeriana, Bogota,  
(Colombia)

E-mail: [yurigor@javeriana.edu.co](mailto:yurigor@javeriana.edu.co)

### **Resumen:**

**Objetivo:** Diseñar un modelo formal de la estructura de incentivos creada por la evaluación del profesor por los estudiantes y hacer su prueba empírica.

**Método:** Un juego dinámico de información perfecta en tres etapas. Para probar el modelo, se utilizan observaciones realizadas en un curso introductorio de administración en 2013 - 2014. Se analiza la correlación entre la nota promedio y la evaluación del profesor por los estudiantes para cada clase de la asignatura.

**Resultados:** El modelo muestra equilibrios múltiples y existencia de la correlación en casi todas las celdas. En algunas celdas la correlación es deseable, en otras, indeseable.

**Conclusiones:** La evaluación del profesor por los estudiantes se correlaciona con las calificaciones de los alumnos lo que conduce a la inflación de notas. El grado de la inflación depende de la política del administrador quien puede prestar una mayor o menor atención a las evaluaciones de profesores por los alumnos.

**Palabras clave:** Educación en administración; incentivos; evaluación del profesor por los alumnos

**Abstract:**

*Objective: To design a game theoretic model of incentive structure created by the student evaluation of teacher and to make an empirical test of the model*

*Method: I use game theory to build a dynamic perfect information game in three stages. To prove the model, I collect observations in the Management introductory course in 2013 - 2014 and make test of correlation between the average grades per class and the student evaluation of teacher.*

*Results: The model shows multiple equilibria and existence of the correlation in almost every cell. In some cells the correlation is desirable, in other cells, undesirable.*

*Conclusions: The student evaluation of teacher is correlated with the grades and results in grade inflation. The degree of the inflation depends on the university manager who pays more or less attention to the student evaluation of teacher.*

*Key Words: Management education; incentives; student evaluation of teacher*

## 1. Introduction

The student is asked to fill the SET evaluation form individually on line. The individual evaluations are aggregated for every class of every course. The final evaluation shows average numbers for the class and the number of students in the class who filled the form, but not their names. The final evaluation is delivered to the management of the academic department and every teacher involved. The evaluation is one of the inputs that inform the decision of the career advancement of full time teachers. The evaluation is a critical input to decide continuity of part time teachers.

The universities managers assume that students are autonomous mature persons, who strive for learning, are capable to tell good teacher from bad one. Additionally, they are truthful and impartial making good evaluation for a good teacher and bad evaluation for a bad teacher regardless of high or low grades obtained in the course.

The assumption is based on a constructivist theory. According to constructivists, the student is not an empty vessel to be filled with knowledge. The student is an active participant in education process. He learns the concepts not by memorizing but when faces problems and tries to solve them. The teacher acts like a guide, facilitator who puts questions, designs exercises and creates adequate conditions for learning. If the student is an active participant, he can evaluate the teacher performance.

The agency theory introduces a little bit of skepticism into this ideal picture. The incentive incorporated in the evaluation, tells the teacher to strive for student satisfaction because a satisfied student will make a good evaluation. What are the components of a student satisfaction? The good grade and the learning (Ewing, 2012), minus the effort needed to get them:

$$U = g + r - e$$

Where:

U is the student satisfaction  
g the grade  
r the learning  
e the effort

What can the teacher do to raise the student satisfaction? First, he can create attractive learning environment and be rigorous in grading. Alternatively, he can streamline the learning process in order to save effort for the student and put high grades. The first strategy is not an easy task. The simplest way to have student satisfied is by means of grades. This is the reason why student evaluation of teacher (SET) is an object of debate. Some

researchers (Johnson, 2003) find correlation between grades and SET and conclude that teachers exchange high grades for high SET scores. Ewing (Ewing, 2012), analyzes the correlation between SET and expected grades. He uses expected, not actual grades because makes the assumption that students do not know their final grade at the time of the evaluation of teachers. He reports positive effect of expected grades on SET. Langbein (Langbein, 2008), Griffin (Griffin, Hilton, Plummer, & Barret, 2014), Braga (Braga, Paccagnella, & Pellizzari, 2014), Schneider (Schneider, 2013) report the same effect. According to the authors, teachers who put their students high grades, rank higher in the SET. Albu (Albu & Badea, 2012) do not report empirical results of the effect of grades on SET but the opinion of the professors who consider that such effect do exist.

Greenwald and Gillmore (Greenwald & Gillmore, 1997) explain the correlation grades - SET with the help of the attribution theory, according to which, students internalize responsibility for their success but externalize responsibility for their failure. When students get high grades, they think that deserve them. When students get low grades, they think that the teacher is responsible. That's why the teacher has the wrong incentive to buy SET scores with high grades.

Other researchers (Marsh & Roche, 2000) think that SET is a good incentive mechanism. They recognize the existence of small (0,2) correlation grades - SET, but do not accept that this correlation may be due to the grade leniency. Marsh proposes an alternative explanation of the correlation which boils down to the following. Good teacher motivates students to work hard. Hard work translates into learning, and learning, into inherent satisfaction and high grades. High student satisfaction and high grades explain high SET. That's why the correlation grades - SET is not dangerous and has nothing to do with grade leniency and grade inflation.

The same fact (correlation between the grades and SET) has two opposite interpretations. Why does this situation happen? First, because the academic community lacks a conceptual model that explains the nature of the incentive structure created by SET for the teacher and the student.

Second, because the empirical evidence about the correlation grades - SET is mixed. Lack of a strong empirical evidence on the existence of correlation grades - SET has to do with the complexity of the phenomena of teaching and learning. It also has to do with the method used by previous researchers. Published papers analyze a range of courses picked up in various academic departments. This variety is a source of noise because the researcher makes comparisons among courses and academic departments which have different appeal, pedagogical method, grading policy and different reputation among the students. Things are getting worse when researchers include in their samples small elective courses composed of one or some classes. In small elective courses students have a strong motivation and know in advance the name of the teacher. These elements make that the assumption of randomness of enrollment does not hold and the possible grade - SET correlation become spurious (Marsh & Roche, 2000). There are some exceptions like Griffin (Griffin, Hilton, Plummer, & Barret, 2014) that do control for academic department, teacher and course.

The aim of this paper is to design a game theoretic model which explains the incentive structure created by SET, and make an empirical test of the model.

My work differs from the previous literature which explores existence of correlation between grades and SET. I try to create a model which may be a missing link to explain the

grade - SET correlation. It makes a practical contribution because helps to shape incentive policy in universities.

My empirical data are collected in one university and in only one course composed by many classes. This limitation to one course is designed to minimize effects of different pedagogical methods used in various academic departments and courses.

My empirical data are collected in a large Colombian private university, and my work is the first study of this kind in Latin America, as far as I know. In a wider sense, my paper also contributes empirical evidence to the pay for performance literature.

## 2. Method

The incentive situation is modelled as a dynamic game of perfect information in three periods. In the first period the Nature decides if the teacher is good or bad. In the second period the teacher selects the strategy of grade leniency or of grade rigor. These strategies were formulated according to Marsh (Marsh & Roche, 2000). In the third period the student decides which role he will play: a pragmatic grade seeker or a motivated learner. The payoff represents the utilities of the teacher and the student. The utility of the teacher is SET, the utility of the student is the grade. The game is solved by the backward induction in order to find the optimal combination of the strategies of the teacher and the student.

After the equilibrium strategies have been analyzed, I make an empirical test of the model looking for possible correlation between the grades and SET. This part of the paper is a field study (McGrath, 1994) that use secondary data, and can be classified as an archival study. This type of study allows making direct observations of a system without any disturbance. One of the setbacks of this method is low generalizability because the sample may not be representative of the population. In my case, the sample is intentional because only the students of management introductory course were observed. Nevertheless, the enrollment of the students into the classes is random, and the assignment of teachers to every class is also random. The assignment of the teachers in multi class courses is not known to the students when they select their courses and classes.

In 2014 I approached the Management Department of the Business School of the large private university in Bogota. I obtained and analyzed data collected in three semesters in 2013 - 2014 in the introductory management course. The introductory management course is open to all the students of the University, and is mandatory for the management and accounting majors.

The introductory course is taught in 11-13 parallel classes of 35 - 40 students. The course is taught by 8 - 10 full and part time teachers who use the same program, text, cases and make common exams.

Following Marsh (Marsh & Roche, 2000) I used a class like a unit of analysis. I obtained the average grade for every class of the course and the SET for every class. Total number of observations in my data base is 38. The analysis consisted in bivariate correlation which was performed on SSPS software.

Following Langbein (Langbein, 2008), I use the actual grades. I consider using actual grades rather than expected grades (Ewing, 2012) because the student evaluation of teachers takes place after the final exams, the same week when the final grades are published.

### 3. Results

The Graphic 1 reproduces the game tree. The game has multiple equilibria: Nash equilibrium may be reached only in repetitive game and mixed strategies. The correlation grade - SET exists almost always, irrespective of the quality of the teacher. Let's look at every cell in the game:

**BLM:** in this cell the reader may note a lack of correlation, which is desirable. Motivated students may have incentive to report low quality of teaching when their teacher is bad.

**BLP:** in this cell there is an undesirable correlation. A bad teacher is lenient. A bad student receives high grade and reciprocates to the teacher with high SET.

**GRM:** in this cell the reader may observe a desirable correlation. Rigorous teacher teaches well. Motivated student learns, gets high grades and puts high SET.

**GRP:** in this cell there is an undesirable correlation. Rigorous teacher puts low grade to the pragmatic student. The student, in retaliation, puts low SET score.

**GLM:** a desirable correlation. Good teacher teaches well and additionally is lenient. He puts high grades to motivated students. Motivated student learns and puts high SET

**GLP:** an undesirable correlation. Good teacher is lenient. He puts high grades to pragmatic students. Pragmatic student is grateful and additionally recognizes the quality of the teacher. He puts high SET scores.

In order to make an empirical test of the model, I look for the correlation between SET scores and the average grades for every class. The Table 1 shows a moderate correlation between the SET and the average grade for every class. The correlation is significant at 5%.

Table 1.  
*The correlation between SET and average grades*

Variables and indicators / Variables		Average grade	SET
Average grade	Pearson correlation	1	,409 <sup>*</sup>
	Sig. (bilateral)		,011
	Number of observations	38	38
SET	Pearson correlation	,409 <sup>*</sup>	1
	Sig. (bilateral)	,011	
	Number of observations	38	38

\* The correlation is significant at 0,05 (2 tales) level

Source: author calculations  
 Program used: Excel 1997

#### 4. Discussion

The SET mechanism does not allow to tell good from bad teacher because the SET scores are difficult to interpret: the SET - grade correlation may signal quite different things. The correlation may signal a good teacher quality, a desirable result suggested by Marsh: good teacher teaches well and the student gets high grades. The correlation may signal the grade leniency, but the correlation cannot make evident if the lenient teacher is good or bad. The lack of correlation may signal two different things. First, a truthful reporting of low teacher quality by a good student. The lack of correlation also may signal a truthful reporting of a high teacher quality by a bad student.

In all these situations the existence of correlation leads to grade inflation. This result is similar to that obtained by Schneider (Schneider, 2013). For illustration, compare the GR and GL strategies. The GR strategy is risky for the professor because he will put low grade for the pragmatic student and runs the risk to get a low SET score in retaliation. To the contrary, the GL strategy does not present such risk: in both cells GLM and GLP the teacher gets high SET score. For a teacher, to be rigorous is a risky business; to be lenient is a safe dominant strategy.

The same business school where the data were collected illustrates the grade inflation. Before 2011 it applied the 3,2 as an average grade necessary to stay at the school. In 2011 this average grade was raised to 3,4 without any effect on quality of graduates and the school ranking.

The absence of equilibrium on pure strategies suggests that teacher and student do not play always the same strategy. They know that they play a repeated game and use mixed strategies: sometimes the teacher is rigorous, sometimes is lenient; sometimes the student appears motivated; sometimes he behaves like a pragmatic. This result is similar to that obtained by Griffin (Griffin, Hilton, Plummer, & Barret, 2014) who found that teachers change their behavior when teaching in a range of classes.

The university administrator helps to shape the strategy mix used by the teacher. If the university manager pays much attention to SET, he will press teachers toward “grade leniency” strategy. If the manager does not pay much attention to SET, he signals that the teacher is free to be demanding with the students.

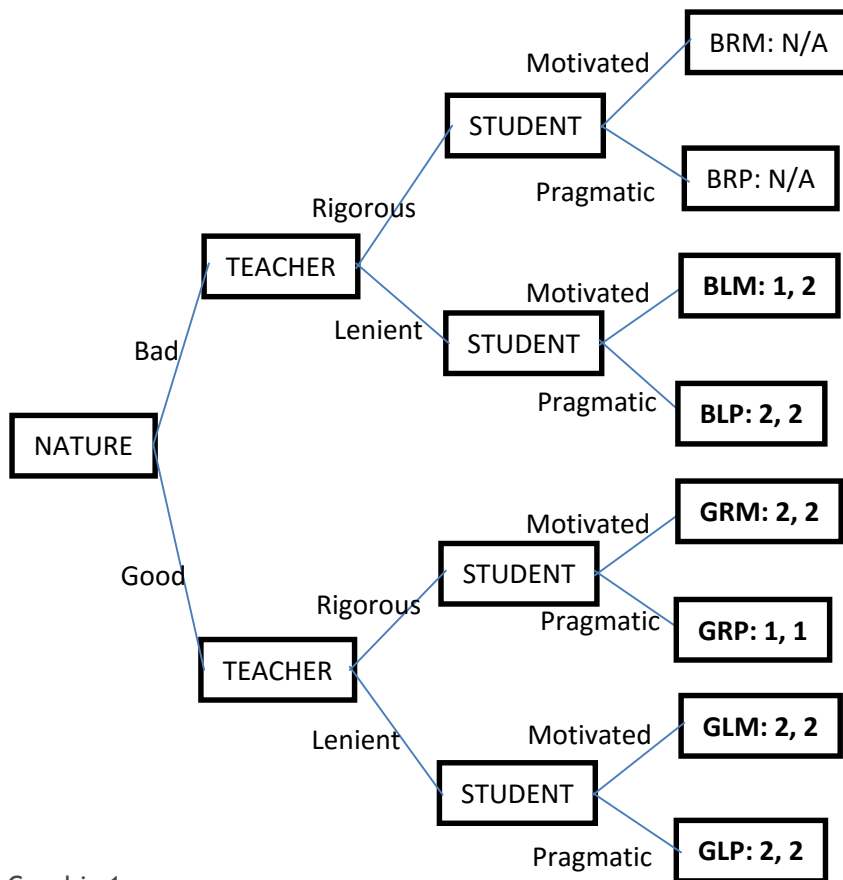
What can be done to solve the incentive problem created by the SET? The game shows that only motivated students have incentive to truly report the teacher quality, while the pragmatic students have no such incentive. This result suggests that one way out is to eliminate the student anonymity from SET. The teacher and the manager must have access not to the aggregated but to the individual SET scores put by every student, and to relate these scores to students’ grades. This strategy will boil down to taking into account only the opinions of motivates students who approve the course. To avoid ethical problems, the students may use pseudonyms and the management may make SET after the final grades are handed over to the students.

My work has some limitations. The simple dynamic one shot game does not represent the repeated nature of teacher - student interaction and does not allow modeling the probability of using certain strategy. Future work in this field may be based on a repeated game where the participants use mixed strategies. May be useful a signal game which allows to detect a true nature of a participant notwithstanding her declarations. As to the empirical

test, my data base has little observations. Future researcher should collect a bigger data base in some separate courses in order to avoid statistical noise.

### Referencias Bibliográficas

- Albu, G., & Badea, M. (2012). The Student - Evaluator of the Teacher. A Case Study in the University of Ploiesti, Romania. *Procedia - Social and Behavioral Sciences*, 69(24), 2257-2263.
- Braga, M., Paccagnella, M., & Pellizzari, M. (2014). Evaluating students' evaluations of professors. *Economics of Education Review*, 41(August ), 71-88.
- Ewing, A. (2012). Estimating the impact of relative expected grade on student evaluations of teachers. *Estimating the impact of relative expected grade on student evaluations of teachers. Economics of Education Review, Volume 31, Issue 1, February 2012, Pages 141-154. Andrew M. Ewing (Science direct Elsevier)*, 31(1), 141-154.
- Greenwald, A., & Gillmore, G. (1997). No pain, no gain? The importance of measuring course workload in student ratings of instruction. *Journal of Educational Psychology*, 89, 743-751.
- Griffin, T., Hilton, J., Plummer, K., & Barret, D. (2014). Correlation between grade point averages and student evaluation of teaching scores: taking a closer look. *Assessment & Evaluation in Higher Education*, 39(3), 339-348.
- Johnson, V. (2003). *Grade inflation: a way out*. Springer.
- Langbein, L. (2008). Management by results: Student evaluation of faculty teaching and the mis-measurement of performance. *Economics of Education Review*, 27(4), 417-428.
- Marsh, H., & Roche, L. (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), 202-228.
- McGrath, J. (1994). *Methodology matters: doing research in the behavioral and social sciences*. Urbana: University of Illinois.
- Schneider, G. (2013). Student Evaluations, Grade Inflation and Pluralistic Teaching: Moving from Customer Satisfaction to Student Learning and Critical Thinking. *Forum for Social Economics*, 42(1), 122-135.



Graphic 1  
The game tree.

Source: author designed  
Program used: Excel 1997