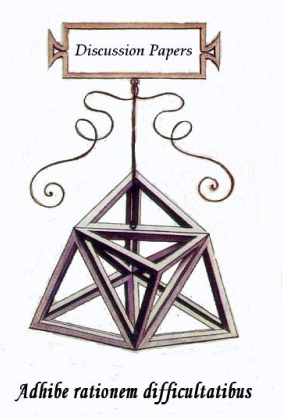
Discussion Papers
Collana di
E-papers del Dipartimento di Economia e Management – Università di Pisa



Angelo Antoci - Irene Brunetti - Pierluigi Sacco - Mauro Sodini

Student Evaluation of Teaching (SET), social influence dynamics, and teachers' choices: An evolutionary model

Discussion Paper n. 225

2017

Discussion Paper n. 225, presentato: **October 2017**

Indirizzo dell'Autore:

Angelo Antoci
Dipartimento di Scienze Economiche e Aziendali, via Torre Tonda 34, 07100 SASSARI – Italy
Email: antoci@uniss.it

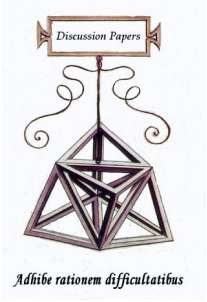
Irene Brunetti
Dipartimento di Economia e Management, via Ridolfi 10, 56100 PISA – Italy
Email: ireneb@ec.unipi.it

Pierluigi Sacco
Dipartimento degli Studi Umanistici, via Carlo Bo 1, 20143 MILANO – Italy
Harvard University and metaLAB (at) Harvard – CAMBRIDGE – United States
Email: pierluigi.sacco@iulm.it

Mauro Sodini
Dipartimento di Economia e Management, via Ridolfi 10, 56100 PISA – Italy
Email: mauro.sodini@unipi.it

Please cite as follows:
Angelo Antoci, Irene Brunetti, Pierluigi Sacco and Mauro Sodini, *(2017) "Student Evaluation of Teaching (SET), social influence dynamics, and teachers' choices: An evolutionary model*", Discussion Papers del Dipartimento di Economia e Management – Università di Pisa, n. 225. (http://www.ec.unipi.it/ricerca/discussion-papers.html).

Discussion Paper
n. 225



Angelo Antoci, Irene Brunetti, Pierluigi Sacco and Mauro Sodini

Student Evaluation of Teaching (SET), social influence dynamics, and teachers' choices: An evolutionary model

Abstract

The issue of student evaluation of teachers (SET) has been explored by a large literature across many decades. However, the role of social influence factors in determining teachers' responses to a given incentive and evaluation framework has been left basically unexplored. This paper makes a first attempt in this vein by considering an evolutionary game-theoretic context where teachers face a two-stage process where their rating depends on both students' evaluation of their course and on retrospective students' evaluation of their teaching output in view of students' performance in a related follow-up course. We find that both high effort (difficult course offered) and low effort (easy course effort) outcomes may emerge, and that may either lead to a socially optimal outcome for teachers or not, according to cases. Moreover, there may be a potential conflict between the optimal outcome for students and for teachers. We also consider possible ways to generalize our model in future research.

Keywords: Student evaluation of teachers (SET); teacher effort; teacher motivation; social selection; strategic interaction between teachers.

JEL: I21, C73, D79.

Student Evaluation of Teaching (SET), social influence dynamics, and teachers' choices: An evolutionary model

1. Introduction

Although employee evaluation is a common practice in work environments (Rynes et al., 2005; Cahuc et al., 2014), in the case of higher education teachers the anomalous aspect is that the evaluation is generally carried out by their students – a peculiarity that raises concerns of validity and reliability (Zhao and Gallant, 2012). In traditional workplace settings, the responsibility of the evaluation is with the worker's supervisor, namely a subject in a higher hierarchical position that is strategically interacting with the worker in a classical principal-agent context (Chauvet et al., 2015). In such context, the evaluation of the worker's performance is an integral part of the principal-agent scheme itself (Mitusch, 2006). In the case of the student evaluation, however, the evaluator is in a strategically subordinate position, although protected by anonymity. Student evaluations of teachers are largely adopted in higher education institutions, and their outcome may have a significant impact on the latter's professional opportunities and even career prospects (Krautmann and Sander, 1999). Student evaluation of teaching (SET) is therefore considered an integral part of the educational and training process, and such evaluations are today the most important, and sometimes the sole, measure of a teacher's ability other than traditional forms of peer evaluation or self-assessment (Greenwood and Ramagli, 1980). This creates a natural incentive for teachers to manipulate the scheme to their own advantage (Roberts, 2016), e.g. by inflating grades to positively influence students' evaluations (Ewing, 2012), with an obvious information bias on both actual quality of teaching and students' performance (Langbein, 2008).

In terms of how SET is carried out and used by school and university deans, however, there is a far from uniform situation, both across different national institutions and a fortiori at an international level. SET is usually administered through anonymous questionnaires filled by students, but its structure, the way in which it is administered and collected, data processing, techniques of analysis and performance indicators, and nature of the feedback to the evaluated teachers may all largely differ from case to case. Deans are the only ones to whom full information about the performance of teachers and the respective performance indicators is disclosed, with an implied large amount of discretional power in how they are interpreted, disseminated, and used in decision-making. As a rule, all questionnaires focus upon basic features of teaching such as clarity, perceived competence, relevance, internal consistency, syllabus appropriateness, quality of teaching materials and reading lists, fair balance between course requirements and credits, and contextual features such as availability to students and punctuality both in class and at office hours, performance of teaching assistants, classroom logistics, etcetera (Braga et al., 2014). Questionnaires may also cover additional features according to cases.

The debate about whether SET is a useful tool for teachers' evaluation or not, and consequently about whether they cause biases in teachers' grading choices, is still open and heated. The literature is not entirely conclusive about the usefulness of SET in the light of the possible incentive compatibility problems that it raises (Darwin, 2017), of the necessity of further methodological development (Setari et al., 2016), and of the ambiguity of the very notion of 'good' teaching from the viewpoint of students (Nasser-Abu Alhija, 2017), which in turn also partially depends on cultural differences (Georgakopoulos and Guerrero, 2010). On the other hand, constructive feedback from student evaluations seems to be helpful in improving teachers' performance (Knol et al., 2013), and teachers' perceived care for students may have a larger positive effect on student evaluation than expected grades (Gotlieb, 2013). The main issue is of course the tradeoff between securing a high quality of teaching vs. manipulating the scheme at the advantage of both parties. In principle, both teachers and students profit from a high-quality teaching environment. Students benefit from the high level of qualification that they acquire through attendance and study and from a higher level of intrinsic motivation and engagement (Griffin, 2016), whereas teachers get the reputational benefits from teaching in an institution that provides excellent education, which may result in further professional opportunities and career advances, plus they enjoy a fulfilling professional experience. On the other hand, there is a clear public good dilemma in that, once

the high reputation of the institution has been established, there is an incentive to free-ride for both teachers and students (Matos-Diaz, 2012), or to set up positive reciprocation schemes (Cho et al., 2015). Students may be tempted to find ways to get high grades while economizing on studying effort (Mangan and Fleck, 2011), whereas teachers may be in turn tempted to receive good evaluations by accommodating the students' shirking attitude through grade inflation, getting higher chances of outcompeting colleagues for tenured positions (Johnson, 2003). If this is the case, the overall performance of the educational institution is compromised, and this will eventually cause a loss in reputation. The SET mechanism, unfortunately, may implicitly set up incentives for both parties to mutually adjust in terms of optimal free-riding, and may even penalize pedagogical innovation (Walder, 2017). In terms of social optimum, as it is typical of public good problems, the high-quality equilibrium is Pareto superior to the low-quality one. However, the outcome of a SET-driven quality monitoring strategy may be Pareto-suboptimal due to the dysfunctional incentive structure, moreover causing a reduction of the signaling value of education for the screening of workers in the labor market. Recent evidence suggests though that only less than half of the increase in average grades over two decades at a large US public university (Clemson) may be attributable to grade inflation factors (Hernández-Julián and Looney, 2016).

The modelling of the interaction between teachers and students in a SET environment naturally lends itself to be deployed in game-theoretic terms, and there is a substantial amount of literature that follows this route. However, relatively little attention has been devoted so far to the social influence dynamics that govern strategic behavior in this context. The extent to which teachers may be prone to inflate their grades, or students to shirk on their performance, may also depend upon social incentives, such as conforming to established collective behaviors. However, the literature so far tends to regard choices on both sides as individual ones, with little attention to the social environment. This paper offers a new contribution to the literature on SET-related strategic interaction that places it in a social context, and where consequently social selection of behaviors takes place. Moreover, our model considers a sequential strategic interaction where teachers' quality choices in a course affect students' performance in a subsequent, related course, as students' performance in the second course is also dependent on the knowledge acquired in the first. Therefore, if most students fail in the second course, this may be seen as an indirect signal of negative quality of the teaching in the first course (although, as shown by Carrell and West, 2010, teaching methods that positively affect students' evaluations of a course might also harm their follow-on achievement in subsequent, more advanced courses). In our model, then, reputation effects for teachers play an important role in their strategic decisions.

We characterize the conditions under which a Pareto efficient outcome where teachers provide quality classes and students work hard and reward teachers with good evaluations emerges as the result of social selection. Depending on cases, the efficient outcome may prevail for all initial distributions of behavioral types across the population of teachers, or it may require that an initial high enough critical threshold of high-performing teachers is found. Intuitively, a critical role in determining these conditions is played by the discount factor: the more teachers keep into account the effect of their teaching performance on the students' preparedness in the subsequent course (and therefore their own future evaluation on the basis of the students' performance in the subsequent course), the more likely they will choose to teach a good course. The more such a forward-looking attitude pays off for teacher, the more it tends to spread socially across the population of teachers, and to become an ingrained feature of the educational environment, and vice versa.

The remainder of the paper is organized as follows. Section 2 presents the literature review. Section 3 introduces the model. Section 4 introduces the social dynamics, illustrates the basic results and develops the welfare analysis. Section 5 provides a final discussion and concludes.

2. The under-recognized social dimension of SET

The literature on SET has a long history, dating back to more than 80 years (Linse, 2017), and an interesting persistence. The contemporary debate is still influenced to some extent by comprehensive assessments from the mid-70s (Page, 1974), and by statistical approaches to the measurement of their effectiveness developed in the early 80s (McCready, 1981). Also the literature on the behavioral implications of SET mechanisms spans several decades. Rotem and Glasman (1979) provide an early warning on the source and nature of bias of SET, and Kroman (1978) underlines how the teacher's and student's perspectives in SET may be both limited in their focus and incapable to take the other side's position into account. Brown and Saks (1987) analyze teachers' time allocation choices and consider how strategic behavior can affect them. Wilson (1998) presents a review of the fundamental critical issues to be tackled by SET designers. In the late 90s, however,

SET has become a fully established practice, with a key role in faculty hiring and promotion decisions (Becker and Watts, 1999).

In the subsequent years, the literature on SET has proliferated to an extent that makes it almost impossible to make a fair appraisal of the pros and cons of its use while taking into account all of the available evidence (Pounder, 2007). Despite the huge research and measurement effort, the literature is therefore still inconclusive. Evaluation studies have not managed so far to bring about a consensus about the effectiveness of SET, and meta-analyses of the literature even suggest that there might be no significant relationship between students' evaluations of a teacher's performance and actual students' learning from that teacher (Clayson, 2009; Uttl et al., 2017). Here, rather than attempting a comprehensive survey of the whole body of literature, which would largely exceed the space limitations and the scope of this paper, we focus on a specific, under-recognized aspect of SET: the role of social incentives in determining both teachers' and students' behavior, in their strategic interplay with the incentives set up by the functioning of the SET mechanism itself. The social dimension plays a truly important part in SET, as the teacher-students relationship takes place in the micro-social environment in the classroom. Moreover, both teachers and students constantly interact with their peers both within the context of their own educational institution and of other, often spatially close, ones. These interactions inevitably influence many different aspects of teachers' and students' behaviors, from role models and perceived social norms, to expectations about incentives and rewards, to expectations about future career prospects, and so on. Therefore, evaluating the effects of SET as abstract mechanisms without keeping into account the specific social conditions in which a given mechanism operates may be misleading. Depending on the social environment, the same mechanism could yield either socially optimal or disappointing results, according to circumstances.

Recent research is starting to reflect these subtleties, although generally without an explicit focus on the role of the social environment. As a rule, relatively more motivated, committed, well-performing students tend to participate to the evaluation process more than other types of students; therefore, where educational systems work well, we expect higher levels of participation to, and possibly a more effective functioning of, SET (Kherfi, 2011). Gaertner (2014) reports for instance the results of a German case study where students provide reliable assessments of teachers' performance, and teachers constructively discuss students' feedback with them and adapt their teaching methods accordingly. However, the extent to which these results may depend on the deeply ingrained cooperative social governance model of German society cannot be ignored (Orrù et al., 1998). Likewise, students at the Belgian University of Antwerp tend to provide better SET evaluations the more they perceive SET to have a value as a tool for improving quality of teaching (Spooren and Christiaens, 2017), implicitly manifesting their reliance on evaluation mechanisms in a social context which has historically been characterized by high levels of formalized social monitoring (Hofman, 2014). On the other hand, in contexts with low social capital and strong reliance on informal ties and familism, such as in Southern Italy, the evaluation of teachers may be less compelling and systematic patterns of grade inflation may be observed (Argentin and Triventi, 2015).

In the existing literature on SET, the role of social incentives pops up often although unsystematically, but generally lacks a clear conceptual framework that highlights the potential connections between different results. One key aspect of traditional, university-administered SET is their confidential character, and the fact that their result is not disclosed to students or peers unless the teacher intentionally does it. Therefore, from the point of view of social influence mechanisms, analysis of publicly available sources of information on teacher evaluation, such as online platforms for the evaluation of teachers like ratemyprofessors.com, may be of special interest, as these platforms provide the basis for 'electronic word-of-mouth communication' (Hartman and Hunt, 2013). In such platforms, students voluntarily post their assessment of a teacher's educational performance, and it turns out that such pooling of information not only impacts other students' expectations about classroom experience and attitude toward the class, but also improves their perceived control, both at the undergraduate (Kowai-Bell et al., 2011) and at master level (Kowai-Bell et al., 2012), thus establishing a powerful channel of social influence where single reviews may acquire a disproportionate weight. Not surprisingly, it is found that the availability of such kind of information tends to influence students' course choices independently of its reliability, and leads to strong biases in choice (Li and Wang, 2013). On the other hand, such online evaluations, despite their well-known limitations in reliability and representativeness, also impact upon teachers' affect and self-efficacy (Boswell, 2016), though not upon their self-concept of competence (Kowai-Bell et al., 2012).

Perhaps more surprisingly, another result that emerges from the literature and points to social influence effects is that SET tend to be sensitive to race and gender (Basow and Silberg, 1987; Bavishi et al., 2010; Basow et al., 2013), and are even systematically influenced by the perceived sexual attractiveness of the teacher (both

male and female) – an aspect which is clearly uncorrelated with teaching performance (Riniolo et al., 2006; Fenton et al., 2008). Wagner et al. (2016) find evidence of a particularly strong negative gender bias against women teachers even in a diverse, multi-ethnic, multicultural sample of students and teachers from a Dutch university, and Boring (2017) finds similar evidence of negative gender bias against women in a French university. A more ambiguous factor is the amount of social interaction between teacher and student, that consistently predicts positive student evaluation and may be partially related to teaching quality, but certainly also accounts for some dimension of social communication and influence (Sheer and Fung, 2007). There is moreover a significant amount of subjective variation in students' relational responsiveness to different teachers, with potential gains from appropriate matches (Gross et al., 2015).
The previous discussion shows that there is a variety of social incentives at work that may influence the functioning of SET in various ways and directions. Therefore, failing to take into account social influence effects may be a major model shortcoming in the attempt to understand under what conditions SET are able to promote the achievement of socially optimal results. We will now present an evolutionary game-theoretic model that provides a simple context that allows us to deal with the social selection of a SET-driven optimal outcome.

3. The model

Several game-theoretic models of the teacher/student and teacher/teacher strategic interaction are currently available. Building on the seminal work of Marchi and Miguel (1974) and Hamburger (1979), Correa and Gruver (1987) model the teacher/student strategic interaction with a continuous strategy set, and find that non-optimal allocations can emerge due to a suboptimal level of effort provided by both teachers and students. However, the introduction of a teacher evaluation system may lead to a higher level of effort than required by the social optimum, possibly leading to dysfunctional over-commitment effects (Reimann, 2016). An early, similar modeling of the teacher/teacher interaction is proposed by McKenzie (1979), who considers the joint offering of a course by two teachers in two distinct modules, with the common aim of attracting the largest possible number of students. Correa (2001) shows that this setting naturally leads to a social dilemma situation with the well-known sub-optimality issues. Correa (2003) analyzes the strategic interaction among one teacher and n students with different abilities and attitudes toward effort, analyzing the incentives for the teacher to provide a more or less committed approach to teaching, and introducing the issue of diversity in players' capabilities and ethical standards. Strategic behavior of teachers is of particular relevance in view of the consolidated evidence that teachers are sensitive to economic incentives (Figlio and Kenny, 2007), and that monetary incentives may crowd out teachers' intrinsic motivation and attitude toward unpaid work (Jones, 2013). In this paper, we combine some of the previous elements by considering a situation where two teachers are called to cooperate in the achievement of high teaching standards having both to choose between two different levels of teaching output. However, we also add a sequential element to our model, namely, that the teacher's output influences the performance of students in a subsequent course, thus introducing a reputational effect that plays against the incentive to free ride on effort.
We postulate in particular that teachers are evaluated in two stages: immediately after the end of their course, and once more at the end of a second, related course that their students attend subsequently. Students' evaluation of the second course also provides an additional source of evaluation of the teachers of the first course, whose total evaluation is a weighted sum of the two. By providing low effort teaching output, teachers of the first course therefore make less likely that the students are well prepared forth second course, and consequently obtain relatively worse evaluations that also negatively affect the teachers of the first course, who may consequently have an extra incentive to provide high effort output ceteris paribus. Will this be enough to ensure that the social optimum is reached?
More formally, teachers face a strategic choice between offering a demanding (Difficult, D) or a less demanding (easy, E) course. We consider a large population of teachers where, at each time t (time is continuous), two of them are matched to play an evaluation game. The payoffs to the strategies D and E are determined through a two-stage evaluation process. The two stages are indicated as I and II, respectively. At stage I, teachers are evaluated by their students. At stage II, they receive a second evaluation as their students, by retrospectively evaluating their performance in a subsequent, related course, rate the first course professor's actual contribution to their preparedness for the second course. As far as the stage I evaluation is concerned, we assume that students prefer professors who give them a relatively light study load and relatively good grades, that is, they prefer to attend an E rather than a D course. Consequently, we assume for simplicity that

at stage *I*, *E* deterministically ensures better student evaluations than *D*, and that such difference in evaluations reflects into teachers' payoffs. On the other hand, we assume that teachers' payoffs are not influenced by socially relevant factors such as gender, ethnicity, or sexual attractiveness. The only social incentives that matter in our model are therefore linked to the frequency of adoption of certain behaviors, but not to players' (teachers') personal traits. In terms of the evaluation game for the two teachers, the (symmetric) payoff matrix for stage *I* is

$$\text{Stage } I: \quad \begin{array}{cc} D & E \\ \alpha & 0 \\ 1 & \beta \end{array} \begin{array}{c} D \\ E \end{array} \qquad (1)$$

where the payoff for the best outcome for teachers has been normalized to 1. For a (row) teacher, the outcome *(E, D)* is the optimal one in that the teacher provides low effort and receives a good evaluation, and in particular a better one than the column teacher providing high effort (Krautmann and Sander, 1999; Oleinik, 2009). Accordingly, the worst outcome, normalized to 0, is the *(D, E)* one where the teacher provides high effort and receives a worse evaluation than the other teacher providing low effort. Consequently we have that $1 > \alpha$ and $\beta > 0$, and thus strategy *E* dominates *D* in the single-stage game. Moreover, if $\alpha > \beta$ the game turns into a Prisoner's Dilemma, with *(E, E)* as the unique Nash equilibrium and *(D, D)* as the social optimum. This payoff structure might hold if teachers, even when sensitive to the strategic temptation to shirk, still maintain some level of intrinsic motivation for teaching quality that makes socially uniform levels of high effort teaching preferable to uniform levels of low effort teaching when no personal strategic advantage may be reaped from the interaction. If we admit moreover the possibility that $\alpha > 1$, so that teachers strongly prefer the uniformly high effort social situation to free riding, then the game admits two Pareto-ranked Nash equilibria. From the students' perspective, *(D, D)* would be preferred to *(E, E)* in that they have an interest in maximizing the knowledge return to their educational investment. In what follows, however, we restrict for simplicity the analysis to the case $\alpha < 1$.

At stage *II*, we assume that *D* deterministically ensures teachers a better evaluation than *E*, and that such difference in evaluation reflects into teachers' payoffs, whereas again other socially relevant individual characteristics such as gender, ethnicity or sexual attractiveness do not matter. The payoff matrix for teachers is then the following:

$$\text{Stage } II: \quad \begin{array}{cc} D & E \\ a & 1 \\ 0 & b \end{array} \begin{array}{c} D \\ E \end{array} \qquad (2)$$

where again the best outcome has been normalized to 1, and the worst to 0. Now, from the second course's perspective it is optimal for the teacher to have provided a high effort teaching output at stage *I*, since this now ensures a positive retrospective evaluation by students after they have taken the second course, and the more so if the other teacher provided a low effort teaching output, with a consequent lower preparation for the students. Here too we have $1 > a, b > 0$, and once again the relative size of *a* and *b* depends on teachers' attitudes toward their teaching duties, and ultimately on their motivations. With respect to stage *I*, the payoff structure is now overturned, and in the single stage game strategy *D* dominates *E* so that, if the strategic interaction were limited to stage *II*, all teachers would choose to provide a high effort teaching output in the first course. Under the postulated payoff structure of the two stages combined, teachers now face a trade-off: getting a better payoff in the short run by playing *E*, or being focused on the long run by playing *D*.

4. Social selection dynamics

4.1 Evolutionary dynamics

Assume that the population of teachers is very large, and that at each time *t* two teachers are randomly matched from the population to play the two-stage evaluation game (1) - (2) introduced in the previous section. In this

context, time t may be interpreted as a parameter that orders the evaluation events. Teachers choose their strategies ex-ante, without knowing the strategy chosen by the other teacher. Denote by $x(t)$ the share of teachers choosing strategy D at time t. Strategy E will be consequently chosen by $1-x(t)$ teachers at t, with $1 \leq x(t) \leq 1$. The population shares of the two strategies also represent, in a random matching environment from a large population, the probabilities to be matched to a teacher choosing the respective strategy. To determine the teachers' payoff over the two stages, we assume that payoffs earned at stage II are weighted (discounted) by a factor $\theta \in [0,1]$, so that the expected payoffs accruing to strategies D and E are given, respectively, by

$$\pi_D(x) = \alpha \cdot x + \theta \cdot [a \cdot x + 1 \cdot (1-x)] = [\alpha - \theta(1-a)]x + \theta \qquad (3)$$

$$\pi_E(x) = 1 \cdot x + \beta \cdot (1-x) + \theta \cdot b \cdot (1-x) = (1 - \beta - \theta b)x + \beta + \theta \cdot b \qquad (4)$$

We model the social selection dynamics for the two strategies in terms of a payoff-monotonic evolutionary dynamics which, in the case of two strategies, may be specified without loss of generality in terms of the replicator dynamics (Weibull, 1995):

$$\dot{x} = x(1-x)[\pi_D(x) - \pi_E(x)] \qquad (5)$$

where $\dot{x}$ is the time derivative of $x(t)$, $\dot{x} = dx(t)/dx$, whereas the payoff differential is given by

$$\pi_D(x) - \pi_E(x) = \theta(1-b) - \beta + [\alpha + \beta - 1 + \theta(a+b-1)]x \qquad (6)$$

As already noticed, the dynamic behavior of the replicator dynamics (5) is qualitatively equivalent here to that of any sign preserving dynamics of the type $\dot{x} = F[\pi_D(x) - \pi_E(x)]$, where F is a differentiable function in the interval (0,1) such that $\dot{x} > 0$ (respectively, < 0 and $= 0$) if $\pi_D(x) - \pi_E(x) > 0$ (respectively, < 0 and $= 0$). Moreover, under every sign preserving dynamics, the following statements are all true: i) A state $\bar{x} \in (0,1)$ is a stationary state of (5) if and only if $\pi_D(\bar{x}) - \pi_E(\bar{x}) = 0$ holds; ii) The states $x = 0$ and $x = 1$ are locally attractive stationary states if and only if, respectively, (D,D) and (E,E) are (strict) Nash equilibria of the two-stage game defined by the payoff matrices (1) and (2).
The social selection dynamics (5) describes a process where teachers are boundedly rational in that at each instant of time only a small fraction of them considers the possibility of revising their strategy, and the higher the payoff differential between the two strategies at that time, the stronger the (smooth) aggregate shift of strategy-revising teachers from the worse performing strategy to the better performing one.

4.2 Dynamic regimes

The dynamic regimes of the social selection dynamics (5) can be classified as follows.

(A) If $\pi_D(x) - \pi_E(x) \geq 0$ (respectively, ≤ 0) for every $x \in [0,1]$, then we shall say that strategy D dominates strategy E (respectively, E dominates D). If D dominates E, then whatever the initial distribution $x(0) \in (0,1)$ of strategies, the trajectory starting from $x(0)$ approaches the attractive stationary state $x = 1$ (where all teachers play D). Vice-versa, if E dominates D, for any interior initial condition $x(0) \in (0,1)$, the trajectory starting from $x(0)$ approaches the attractive stationary state $x = 0$ (where all teachers play E).

(B) If there exists a repulsive interior stationary state $\bar{x} \in (0,1)$ (where both strategies coexist), separating the attraction basins of the attractive stationary states $x = 1$ and $x = 0$, then we shall say that a *bistable dynamic regime* occurs.

(C) If there exists a stationary state $\bar{x} \in (0,1)$ (where both strategies coexist) and, for any initial distribution $x(0) \in (0,1)$ of strategies, the trajectory starting from $x(0)$ approaches it, then we shall say that a *coexistence dynamic regime* occurs.

Note that:

$$\frac{d\left(\pi_D(x) - \pi_E(x)\right)}{dx} = \alpha + \beta - 1 + \theta(a + b - 1)$$

Consequently, the payoff differential is an increasing function of x if:

$$\alpha + \beta - 1 + \theta(a + b - 1) > 0 \qquad (7)$$

If condition (7) holds, then the relative performance of strategy *D*, with respect to strategy *E*, improves as the share of teachers adopting *D* increases, and vice-versa if (7) is strictly violated. The context in which (7) is *not* met favours the coexistence between teachers playing different strategies, whereas when it is met the extinction of one strategy is generically observed. Essentially, (7) describes how the social incentives at work tend to depend on the aggregate distribution of behaviours across teachers. A situation where the payoff differential between *D* and *E* increases with the share of teachers adopting *D* (and accordingly decreases with the share of teachers adopting *D*), we have a 'snowball' social selection dynamics where the behaviour that becomes socially prevailing eventually takes over at the expense of the other. When on the contrary the payoff differential between *D* and *E* decreases with the share of teachers adopting *D*, we have a 'homeostatic' social selection dynamics that tends to preserve diversity of behaviours across the population of teachers and to reduce the relative share of a certain behavioural type if it becomes too prevalent. As already remarked, the social selection is entirely governed here by the frequency of adoption of the available behaviours, and not by the individual characteristics of the players (teachers). In particular, this also means that the individual characteristics of the teachers make no difference in terms of social salience of their choice from the point of view of the adoption or diffusion dynamics. It is possible to imagine alternative social selection dynamics where this symmetry is violated and the adoption dynamics is biased by factors such as gender, ethnicity, sexual attractiveness, etcetera.
The dynamic regimes of the model can be conveniently described by means of Propositions 1-2:

Proposition 1 *If condition (7) holds, then:*

(*i*) *Strategy D dominates strategy E if:*

$$\pi_D(0) - \pi_E(0) = \theta(1 - b) - \beta \geq 0 \ (i.e.\ \theta \geq \beta/(1 - b)) \qquad (8)$$

(ii) *Strategy E dominates strategy D if:*

$$\pi_D(1) - \pi_E(1) = \alpha + \theta a - 1 \leq 0 \ (i.e.\ \theta \leq (1 - \alpha)/a) \qquad (9)$$

(*iii*) *The bistable dynamic regime is observed if:*

$$\frac{\beta}{1 - b} > \theta > \frac{1 - \alpha}{a} \qquad (10)$$

In other words, *D* dominance requires low discounting of teachers' evaluation at stage II, whereas *E* dominance requires that stage II performance is highly discounted. The bistable regime prevails for intermediate values of θ

Proposition 2 *If condition (7) is strictly violated, then:*

(*iv*) *Strategy D dominates strategy E if:*

$$\pi_D(1) - \pi_E(1) = \alpha + \theta a - 1 \geq 0 \ (i.e.\ \theta \geq (1 - \alpha)/a) \qquad (11)$$

(v) *Strategy E dominates strategy D if:*

$$\pi_D(0) - \pi_E(0) = \theta(1-b) - \beta \leq 0 \ (i.e. \ \theta \leq \beta/(1-b)) \quad (12)$$

(*vi*) *The coexistence dynamic regime is observed if:*

$$\frac{1-\alpha}{a} > \theta > \frac{\beta}{1-b} \quad (13)$$

The intuition for the conditions on θ for proposition 2 in terms of strategy dominance vs. emergence of the coexistence regime follows the same logic highlighted for Proposition 1.
Proofs of Propositions 1-2 are straightforward, and the various dynamic regimes are illustrated by Figures 1-3. For cases *iii* and *vi*, the value of the internal stationary state $\bar{x} \in (0,1)$ is given by

$$\bar{x} = \frac{\beta - \theta(1-b)}{\alpha + \beta - 1 + \theta(a+b-1)} \quad (14)$$

Note that, in the *bistable dynamic regime*, the point $\bar{x}$ is the separator between the basin of attraction of the stationary state $x=1$ (the interval $(\bar{x},1]$) and the stationary state $x=0$ (the interval $[0,\bar{x})$). If the value of $\bar{x}$ increases, then $(\bar{x},1]$ expands while $[0,\bar{x})$ shrinks. The following proposition shows how the value of $\bar{x}$ varies in response to a variation in the discount parameter θ, which of special interest in the interpretation of our results:

Proposition 3

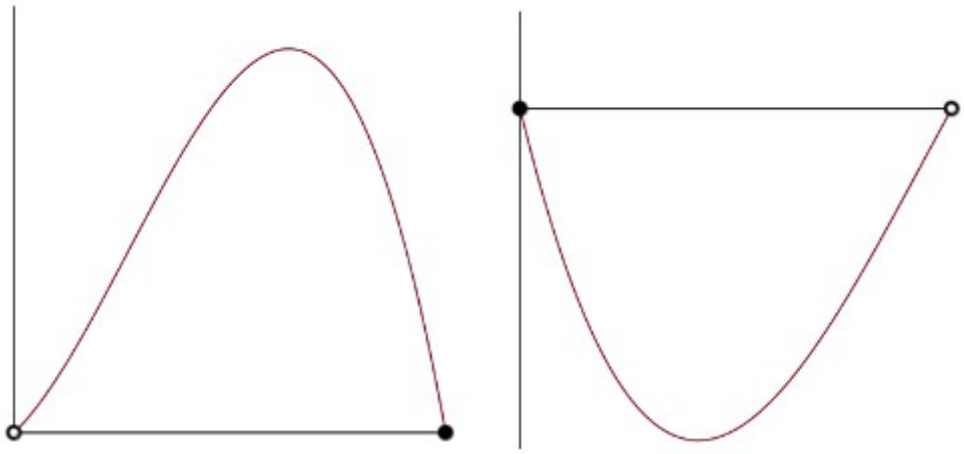
$$sign\frac{\partial \bar{x}}{\partial \theta} = sign(1 + \alpha b - b - \alpha - \beta a) \quad (15)$$

where $1 + \alpha b - b - \alpha - \beta a < 0$ *(respectively,* > 0*) in the bistable dynamic regime (respectively, in the coexistence dynamic regime).*

Proof of Proposition 3 is straightforward. As a consequence of (15), we have that

1) In the *bistable dynamic regime*, an exogenous increase in θ expands the basin of attraction of the stationary state $x=1$ (where all teachers play *D*) at the expenses of the basin of the stationary state $x=0$; this implies that, assuming than the initial strategy distribution $x(0)$ is randomly determined, an exogenous increase in θ has the effect to increase the probability that the state $x=1$ is eventually reached (i.e. that strategy *D* takes over).
2) In the *coexistence dynamic regime*, an exogenous increase in θ has the effect to increase the share of teachers playing *D* at the globally attractive stationary state $\bar{x}$ (i.e. that in the equilibrium mix of behaviors high effort teachers are more represented).

The discount factor θ plays here an intuitively plausible role. The larger the weight that teachers place on the evaluation of their teaching performance at stage *II* (i.e., the less they discount future evaluation at the moment of choosing their strategy at stage *I*), the more strategy *D* will be represented at equilibrium. In particular, it will eventually take over if the social selection dynamics is of the 'snowball' type, or it will be increasingly represented at the equilibrium if the socials election dynamics is 'homeostatic'. All policy measures that will make the follow-up evaluation more salient for teachers, by consequently influencing the size of θ, will therefore prompt a higher incidence of high effort performances across the population of teachers.



(a) (b)

Figure 1. (a) Takeover of high effort strategy *D*. (b) Takeover of low effort strategy *E*.

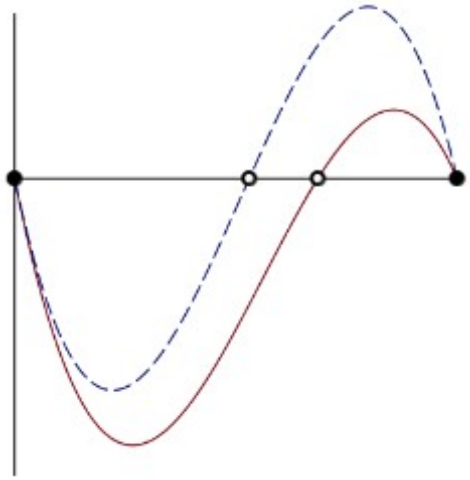


Figure 2. Bistable regime.

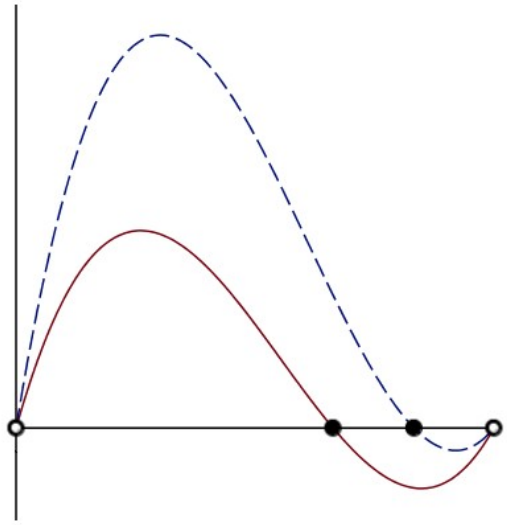


Figure 3. Coexistence regime.

4.3 Welfare analysis

In evaluating the welfare implications of our results, we assume that, for students, a high effort performance of teachers is always preferable to a low effort one, in that students are interested in maximizing the return of their educational investment (Catsiapis, 1987; Levin, 1989; Sun, 1998). Therefore, from the viewpoint of students, the higher the share of strategy *D* at equilibrium, the better off the students. From the point of view of teachers, however, the welfare implications are less straightforward. If view of the payoff structure (1) - (2), teachers' payoffs evaluated at the stationary states 0, 1 and $\bar{x}$ are respectively given by

$$\pi_E(0) = \beta + \theta b$$

$$\pi_D(1) = \alpha + \theta a$$

$$\pi_D(\bar{x}) = \pi_E(\bar{x}) = [\alpha - \theta(1-a)]\bar{x} + \theta$$

It is easy to prove the following

Proposition 4 *In the bistable dynamic regime, where* $x = 0$ *and* $x = 1$ *are both attractive, condition* $\pi_D(1) > \pi_E(0)$ *holds if:*

$$\alpha - \beta > \theta(b-a) \tag{16}$$

In the coexistence dynamic regime, where $\bar{x}$ *is globally attractive, condition* $\pi_D(1) > \pi_D(\bar{x}) = \pi_E(\bar{x})$ *holds if:*

$$\alpha(1-a) > \theta \tag{17}$$

To understand the meaning of Proposition 4, let us consider the bistable dynamic regime and let's assume, to fix ideas, that $\alpha > \beta$ i.e. that the stage *I* game is a prisoner's dilemma where teachers would prefer a uniformly high level of effort, but due to the benefits of free riding the uniform low effort state is the only Nash equilibrium. In this case, $\alpha - \beta$ is the welfare gain for each teacher from achieving the social optimum instead of the Nash equilibrium, that is, the negative of the welfare loss at the Nash equilibrium. If teachers also maintain the same preferences at stage *II*, that is, if they still prefer a uniform high effort state to a low effort one from the point of view of the students' performance in the follow-up exam – that is to say, if they are rewarded enough because of the future positive performance of their students to prefer to exert high effort provided that they cannot benefit from free-riding – then (17) is trivially satisfied, and this means that the high effort equilibrium is welfare improving upon the low effort one in the bistable regime. In this case (which we could call the *goodwill scenario*), therefore, if the initial share of teachers choosing *D* is too small, the social dynamics eventually selects the Pareto inferior equilibrium. If on the contrary teachers' rewards are not strongly dependent on their students' performance at the follow-up exam (despite still preferring to uniformly exert high effort when teaching at their own course rather than uniformly low effort, a case that we could term the *direct responsibility scenario*), so that they like uniformly low effort better from the perspective of the follow-up exam and $b>a$, then the high effort equilibrium will be Pareto optimal only if the stage *II* welfare gain from low effort is discounted enough by teachers, or if alternatively such gain is however smaller than the stage *I* gain from providing uniformly high effort even when the stage *II* gain is not discounted at all. Alternatively, if teachers always prefer the uniformly low effort equilibrium both from the perspective of stage *I* and *II* (and thus in particular $\alpha < \beta$), most likely because their students' performance in the follow-up exam is not imputed to them in that educational environment (a case that we could term the *shirking scenario*), then condition (16) is trivially violated and in the bistable regime the low equilibrium effort is always Pareto optimal, thus creating a trade-off between the welfare benefit for teachers and that for students. In this case, therefore, an excessive initial share of teachers choosing the high effort strategy leads to a Pareto inferior outcome for teachers (but at the same time to an optimal outcome for students) – and this explains why in regimes where shirking-on-the-job social norms prevail, people providing high effort tend to be sanctioned or ostracized by low effort providing peers (Kitts, 2006). Finally, in the case where teachers prefer a uniform low effort state from the point of view of stage I but prefer a uniform high effort state from the point of view of stage II (e.g. because despite their low commitment to effortful teaching they either get monetary or career

benefits if their students do well in the follow-up course, a case that we could term the *instrumentalist scenario*), the welfare comparison between the two equilibria will depend again upon the comparison between the sizes of the welfare loss from a uniform high effort state at stage *I* and the (discounted) welfare gain from a uniform high effort state at stage *II*. In this case, the high effort equilibrium will be Pareto optimal only if the discounted welfare gain from the high effort uniform state at stage *II* will be large enough compared to the welfare loss from the same stage at stage *I*.

In the coexistence dynamic regime, instead, all that matters for the welfare evaluation are the relative sizes of the payoffs at the uniform high effort state in the two stages, and the size of the discount factor. Here, we will always observe a coexistence of the two strategies at the equilibrium, and therefore the initial distribution of types does not have implications for the optimality of the equilibrium state, provided that it lies in the interior of the state space. In this case, then, the high effort equilibrium may only be selected if all players choose the high effort strategy *D* from the beginning.

According to (17), the higher the payoff from the uniform high effort state at stage *I*, and the lower the payoff from the uniform high effort state at stage *II*, the more likely that condition (17) is met. Moreover, the smaller θ, i.e. the more the payoff at stage *II* is discounted, the more likely that the condition is met. The case in which the uniform high effort equilibrium is Pareto superior to the mixed one in the coexistence dynamic regime is therefore one where teachers enjoy a uniform high effort state at stage *I* but not benefit so much from uniformly high effort from the perspective of stage *II*, i.e. they are not rewarded for their students' performance in the follow-up exam but are nevertheless motivated to provide effort in the course they teach. Unlike the analogous case in the bistable dynamic regime, however, here it is not important to what extent teachers enjoy the uniform low effort state, and therefore the emphasis shifts toward the *intrinsic* fact of providing high effort from the viewpoint of stage *I* vs its relative irrelevance at stage *II*. When stage *I* (i.e., when teachers give their own course) matters most in terms of teachers' benefits, we could speak of a *motivation scenario*. If their motivation is strong enough, teachers prefer the high effort equilibrium to the mixed one, and therefore it is enough that any small fraction of teachers initially fails to be motivated by providing low effort to cause a *general* welfare loss. On the contrary, if teachers are not motivated about their effort at stage *I* but are strongly pressured about their performance from the stage *II* perspective, i.e. they are evaluated for the students' final performance and condition (17) is not met, a scenario that we could call the *consequentialist* one, it turns out that the mixed equilibrium is Pareto superior to the high effort one, so that a fair share of teachers may take advantage from the current benefit of slacking provided that the teachers' future benefits from the students' good performance are discounted enough. In this case, the fact that even a small fraction of teachers chooses to shirk causes a *general* welfare improvement.

Notice moreover how conditions (16) - (17) are compatible with conditions (10) - (13), which identify the bistable and coexistence dynamic regimes, respectively, but are not implied by them. This implies that, as we have seen, from the viewpoint of teachers the convergence to either the high effort or the low effort equilibrium may be Pareto optimal in the bistable regime, according to cases, and that, analogously, either convergence to the mixed equilibrium or permanence in the high effort equilibrium may be Pareto optimal in the coexistence regime, according to cases.

5. Conclusions

This paper provides a first analysis of the role of social incentives in determining the effectiveness of SET in the implementation of high effort social standards of teaching. When SET is compelling enough to make teachers accountable for their students' performance in the follow-up exam, the temptation to free ride by getting high scores while offering a low effort course may be overcome in principle but, depending on the details of the incentive structure, this might only happen when an initial critical mass of teachers are willing to provide high effort from the beginning, so that the possibility of being matched to a free riding, low effort teacher is relatively small. If teachers' SET-driven accountability in the follow-up exam is strong enough, however, the high effort equilibrium might prevail eventually even if the initial share of free riders is disproportionately high. Clearly, however, the viability of a strict SET enforcement in the presence of a very high share of low effort teachers may be critical in social and political terms.

Our analysis clearly shows how the best ally of SET in fostering the emergence of high effort equilibria that best serve the interests of students is the teachers' intrinsic motivation to provide high effort, an attitude that reduces the benefit to free ride by providing a low effort course, and consequently making the incentive from teachers' accountability for the students' performance in the follow-up exam more effective. Social incentives

may therefore play a major role in the broader context of SET-driven incentive structures for teachers. On the other hand, the analysis also shows that there may be a conflict between the interest of teachers and that of students as far as welfare considerations are concerned, and the socially optimal outcome for teacher may be one where students do not maximize their educational investment. One might argue that, if students have an objective interest in teachers to provide high effort courses, they should not reward teachers who give a low effort course with better evaluations. However, this remark does not consider the fact that student preferences may be time inconsistent: in the immediate, they tend to prefer an easy pass to a difficult one in a given exam because of limited time resources and pressing deadlines (Zelby, 1974; Brodie, 1998), even if they may be seriously concerned about their educational investment (Entwistle et al., 1974), and the complexity and extent of such inconsistencies substantially depend on different possible learning styles (Entwistle et al., 1979).

In our model, different levels of teacher effort may coexist, or one given attitude may take over, depending on parameter values, and in particular on the motivation and discount rates of teachers (that is, on individual characteristics) and on the 'extrinsic' reward to high-effort teaching on the basis of students' performance in the follow-up task (that is, on systemic characteristics). However, as we have seen, such individual characteristics may lead to different social outcomes, either optimal or not, depending on the social dynamics, and in particular on the initial distribution of behavioral types in the bistable dynamic regime. This result underlines the role of cultural 'contextual' factors, i.e. of the cultural salience of certain behaviors. It may therefore happen that, in regions or countries where established social conventions lead teachers to focus on high effort behaviors, the eventual outcome of the social selection may be opposite to that of other regions where the ruling conventions make low effort teaching salient, despite the fact that both the underlying individual characteristics and the systemic characteristics may in fact be identical. The role of 'critical mass effects' in social selection processes must therefore be carefully evaluated from the viewpoint of policy design. Sometimes, acting on established cultural conventions and social norms may be more effective in terms of the aggregate outcome than manipulating policy parameters or regulating teachers' behaviors through specific evaluation mechanisms such as SET.

Our model presents the simplest possible version of a social selection dynamics of teachers' choices, but clearly more complex models in which socially relevant factors such as gender or ethnicity or personal attractiveness matter, both in terms of students' evaluation and of demonstrative value of teachers' choice at the social level. It would be particularly interesting to study how the selection dynamics operate on social networks with specific relational structures and significant anisotropy in the social interaction patterns. Also, it would be interesting to study models where students with different learning styles, educational investment modes and intertemporal preferences evaluate teachers with different propensities to effort, so that the distribution of teachers' and students' attitudes in the respective populations co-evolve. Evaluating the impact and welfare properties of SET is a rich theme, that lends itself to multiple generalizations with substantial interest both at the theoretical and at the policy level. Our goal in the present paper was to illustrate how such developments appear particularly promising in the so far unexplored dimension of the social selection of teachers' attitudes. Now that the point has been made, we look forward to more research that explore this promising path in its full potential.

References

Argentin G, Triventi M (2015) The North-South divide in school grading standards: New evidence from the national assessments of the Italian student population. Italian Journal of Sociology of Education 7, 157-185.

Basow SA, Codos S, Martin JL (2013) The effects of professors' race and gender on student evaluation and performance. College Student Journal 47, 352-363.

Basow SA, Silberg NT (1987) Student evaluations of college professors: Are female and male professors rated differently? Journal of Educational Psychology 79, 308-314.

Bavishi A, Madera JM, Hebl MR (2010) The effect of professor ethnicity and gender on student evaluations: Judged before met. Journal of Diversity in Higher Education 3, 245-256.

Becker WE, Watts M (1999) How Departments of Economics evaluate teaching. American Economic Review 89, 344-349.

Boring A (2017) Gender biases in student evaluations of teaching. Journal of Public Economics 145, 27-41.

Boswell SS (2016) Ratemyprofessors is hogwash (but I care): Effects of Ratemyprofessors and university-administered teaching evaluations on professors. Computers in Human Behavior 56, 155-162.

Braga M, Paccagnella M, Pellizzari M (2014) Evaluating students' evaluations of professors. Economics of Education Review 41, 71-88.
Brodie DA (1998) Do students report that easy professors are excellent eachers? Canadian Journal of Higher Education 28, 1-20.
Brown BW, Saks DH (1987) The microeconomics of the allocation of teachers' time and student learning. Economics of Education Review 6, 319-332.
Cahuc P, Carcillo P, Zylberberg A (2014) Labor economics. MIT Press, Cambridge MA.
Carrell SE, West JE (2010) Does professor quality matter? Evidence from random assignment of students to professors. Journal of Political Economy 118, 409-432.
Catsiapis G (1987) A model of educational investment decisions. Review of Economics and Statistics 69, 33-41.
Chauvet L, Collier P, Fuster A (2015) Supervision and project performance: A principal-agent approach. UMR-DIAL Document de Travail DT/2015-04, Université Paris-Dauphine.
Cho D, Baek W, Cho J (2015) Why do good performing students highly rate their instructors? Evidence from a natural experiment. Economics of Education Review 49, 172-179.
Clayson DE (2009) Student evaluation of teaching: Are they related to what students learn? A meta-analysis and review of the literature. Journal of Marketing Education 31, 16-30.
Correa H, Gruver GW (1987) Teacher-student interaction: A game theoretic extension of the economic theory of education. Mathematical Social Sciences 13, 19-47.
Correa H (2001) A game theoretic analysis of faculty competition and academic standards. Higher Education Policy 14, 175-182.
Correa H (2003) A game theoretical analysis of educational impact of differences in the abilities and work ethics of teachers and students. Journal of Socio-Economics 32, 249-263.
Darwin S (2017) What contemporary work are student ratings actually doing in higher education? Studies in Educational Evaluation 54, 13-21.
Entwistle NJ, Thompson J, Wilson JD (1974) Motivation and study habits. Higher Education 3, 379-396.
Entwistle N, Hanley M, Hounsell D (1979) Identifying distinctive approaches to studying. Higher Education 8, 365-380.
Ewing AM (2012) Estimating the impact of relative expected grade on student evaluation of teachers. Economics of Education Review 31, 141-154.
Fenton J, Koper PT, Mitchell J, Stinson M (2008) Attractiveness, easiness and other issues: Student evaluations of professors on Ratemyprofessors.com. Assessment & Evaluation in Higher Education 33, 45-61.
Figlio DN, Kenny NW (2007) Individual teacher incentives and student performance. Journal of Public Economics 91, 901-914.
Gaertner H (2014) Effects of student feedback as a method of self-evaluating the quality of teaching. Studies in Educational Evaluation 42, 91-99.
Georgakopoulos A, Guerrero LK (2010) Student perceptions of teachers' nonverbal and verbal communication: A comparison of best and worst professors across six cultures. International Education Studies 3, 3-16.
Gotlieb J (2013) Variables that can affect student ratings of their professors. Journal of Learning in Higher Education 9, 19-27.
Greenwood GE, Ramagli HJJr (1980) Alternatives to student ratings of college teaching. The Journal of Higher Education 51, 673-684.
Griffin BW (2016) Perceived autonomy support, intrinsic motivation, and student ratings of instruction. Studies in Educational Evaluation 51, 116-125.
Gross J, Lakey B, Lucas JL, LaCross R, Plotkowski AR, Winegard B (2015) Forecasting the student-professor matches that result in unusually effective teaching. British Journal of Educational Psychology 85, 19-32.
Hamburger H (1979) Games as models of social phenomena. W.H. Freeman, New York.
Hartman KB, Hunt JB (2013) What ratemyprofessors.com reveals about how and why students evaluate their professors: A glimpse into the student mind-set. Marketing Education Review 23, 151-161.
Hernández-Julián R, Looney A (2016) Measuring inflation in grades: An application of price indexing to undergraduate grades. Economics of Education Review 55, 220-232.
Hofman E (2014) An obligation of conscience: Gossip as social control in an Eighteen-century Flemish town. European Review of History 21, 653-670.
Johnson VE (2003) Grade inflation. A crisis in college education. Springer, New York.

Jones MD (2013) Teacher behavior under performance pay incentives. Economics of Education Review 37, 148-164.
Kherfi S (2011) Whose opinion is it anyway? Determinants of participation in student evaluation of teaching. Journal of Economic Education 42, 19-30.
Kitts JA (2006) Collective action, rival incentives, and the emergence of antisocial norms. American Sociological Review 71, 235-259.
Knol MH, in't Veld R, Vorst HCM, van Driel JH, Mellenbergh GJ (2013) Experimental effects of student evaluations coupled with collaborative consultation on college professors' instructional skills. Research in Higher Education 54, 825-850.
Kowai-Bell N, Guadagno RE, Little T, Preiss N, Hensley R (2011) Rate my expectations: How online evaluations of professors impact students' perceived control. Computers in Human Behavior 27, 1862-1867.
Kowai-Bell N, Guadagno RE, Little TE, Ballew JL (2012) Professors are people too: The impact of informal evaluations of professors on students and professors. Social Psychology of Education 15, 337-351.
Krautmann AC, Sander W (1999) Grades and students' evaluation of teachers. Economics of Education Review 18, 59-63.
Kroman N (1978) Student evaluation of professors: Critique and proposal. Journal of Educational Thought 12, 205-210.
Langbein L (2008) Management by results: Students evaluation of faculty teaching and the mis-measurement of performance. Economics of Education Review 27, 417-428.
Levin HM (1989) Economics of investment in educationally disadvantaged students. American Economic Review Papers and Proceedings 79, 52-56.
Li C, Wang X (2013) The power of eWOM: A re-examination of online student evaluations of their professors. Computers in Human Behavior 29, 1350-1357.
Linse AR (2017) Interpreting and using student ratings data: Guidance for faculty serving as administrators and on evaluation committees. Studies in educational Evaluation 54, 94-106.
Mangan MA, Fleck B (2011) Online student evaluation of teaching: Will professor 'hot and easy' win the day? Journal of Excellence in College Teaching 22, 59-84.
Marchi E, Miguel O (1974) On the structure of the teaching-learning interactive process. International Journal of Game Theory 3, 83-99.
Matos-Diaz H (2012) Student evaluation of teaching, formulation of grade expectations, and instructor choice: Explorations with random-effects ordered probability models. Eastern Economic Journal 38, 296-309.
McCready DJ (1981) Student evaluation of teaching. The Canadian Journal of Higher Education 11, 67-77.
McKenzie RB (1979) The political economy of the educational process. Martinus Nijhoff Publishing, Boston.
Mitusch K (2006) Non-commitment in performance evaluation and the problem of information distortions. Journal of Economic Behavior and Organization 60, 507-525.
Nasser-Abu Alhija F (2017) Teaching in higher education: Good teaching through students' lens. Studies in Educational Evaluation 54, 4-12.
Oleinik A (2009) Does education corrupt? Theories of grade inflation. Educational Research Review 4, 156-164.
Orrù M, Biggart NW, Hamilton GG (1998) The economic organization of East Asian capitalism. Sage, London.
Page CF (1974) Student evaluation of teaching: The American experience. Society for Research into Higher Education, London.
Pounder JS (2007) is student evaluation of teaching worthwhile? An analytical framework for answering the question. Quality Assurance in Education 15, 178-191.
Reimann M (2016) The moderating role of overcommitment in the relationship between psychological contract breach and employee mental health. Journal of Occupational Health 58, 425-433.
Riniolo TC, Johnson KC, Sherman TR, Misso JA (2006) Hot or not: Do professors perceived as physically attractive receive higher student evaluations? Journal of General Psychology 133, 19-35.
Roberts RC (2016) Are some of the things faculty do to maximize their student evaluations of teachers scores unethical? Journal of Academic Ethics 14, 133-148.
Rotem A, Glasman NS (1979) On the effectiveness of students' evaluative feedback to university instructors. Review of Educational Research 49, 497-511.
Rynes SL, Gerhard B, Parks L (2005) Personnel psychology: Performance evaluation and pay for performance. Annual Review of Psychology 56, 571-600.

Setari AP, Lee J, Bradley KD (2016) A psychometric approach to the validation of a student evaluation of teaching instrument. Studies in Educational Evaluation 51, 77-87.
Sheer VC, Fung TK (2007) Can email communication enhance professor-student relationship and student evaluation of professor? Some empirical evidence. Journal of Educational Computing Research 37, 289-306.
Spooren P, Christiaens W (2017) I liked your course because I believed in (the power of) students evaluations of teaching (SET). Students' perceptions of a teaching evaluation process and their relationships with SET scores. Studies in Educational Evaluation 54, 43-49.
Sun Y (1998) The academic success of East-Asian-American students: An investment model. Social Science Research 27, 432-456.
Uttl B, White CA, Wong Gonzalez (2017) Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. Studies in educational Evaluation 54, 22-42.
Wagner N, Rieger M, Voorvelt K (2016) Gender, ethnicity and teacher evaluations: evidence from mixed teaching teams. Economics of Education Review 54, 79-94.
Walder AM (2017) pedagogical innovation in Canadian higher education: Professors' perspectives on its effects on teaching and learning. Studies in Educational Evaluation 54, 71-82.
Weibull J (1995) Evolutionary game theory. MIT Press, Cambridge MA.
Wilson R (1998) New research casts doubt on value of Student Evaluation of Professors. Chronicle of Higher Education 44 (19), A12-14.
Zelby LW (1974) Student-faculty evaluation. Science 183, 1267-1270.
Zhao J, Gallant DJ (2012) Student evaluation of instruction in higher education: Exploring issues of validity and reliability. Assessment & Evaluation in Higher Education 37, 227-235.

Discussion Papers
Collana del Dipartimento di Economia e Management, Università di Pisa

Email della redazione: lfanti@ec.unipi.it