



Student evaluation of teaching, social influence dynamics, and teachers' choices: An evolutionary model

Angelo Antoci¹ · Irene Brunetti² · Pierluigi Sacco^{3,4,5} · Mauro Sodini⁶

Published online: 8 October 2020
© The Author(s) 2020

Abstract

The issue of Student Evaluation of Teaching 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 in which 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 offered) outcomes may emerge, leading either 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 Teaching (SET) · Teacher effort · Teacher motivation · Strategic interaction between teachers · Human capital

JEL Classification I21 · C73 · D79

The authors would like to express their sincere appreciation to an anonymous reviewer for her/his very useful suggestions.

✉ Angelo Antoci
antoci@uniss.it

- ¹ Department of Economics and Business, University of Sassari, Via Muroni, 25, 07100 Sassari, Italy
- ² INAPP, National Institute for Public Policy Analysis, Rome, Italy
- ³ Department of Humanities, IULM University Milan, Milan, Italy
- ⁴ Bruno Kessler Foundation, Trento, Italy
- ⁵ metaLAB (at) Harvard, Cambridge MA, USA
- ⁶ Department of Economics and Management, University of Pisa, Pisa, Italy

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 who 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 the nature of the feedback to the evaluated teachers may all largely differ from case to case. In most cases, deans are the only ones to whom full information about the performance of teachers and the respective performance indicators is disclosed (apart from the confidential performance report received by each teacher on his/her own courses), with an implied large amount of discretionary power as to how they are interpreted, circulated (or purposefully leaked), 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).

The debate on whether SET is a useful tool for teachers' evaluation or not, and consequently on whether they are a source of 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 M.H. 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 professional qualification they acquire through attendance and study, and from a higher level of intrinsic motivation and engagement (Griffin 2016). Teachers get the reputational benefits from teaching in an institution that provides excellent education, which may result in further professional opportunities and career advances, and moreover 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 out-competing 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 individually rational 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 may be 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-up 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 teachers, 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. However, the welfare evaluation of the possible states that are socially selected is complex, and depends among other factors on teachers' motivation and on the relative benefits of high-quality vs. low-quality teaching. As a consequence, the prevalence of high-quality teaching is not necessarily the socially optimal outcome in all circumstances.

The remainder of the paper is organized as follows. Section 2 presents the literature review. Section 3 introduces the model. Section 4 studies the social dynamics and illustrates the basic results. Section 5 develops the welfare analysis. Section 6 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 the bias that characterizes 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 properly 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. By the late 90s, however,

SET had 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 still inconclusive, and evaluation studies have not managed so far to bring about consensus about the effectiveness of SET. The available evidence leaves ample room for concern. Stark and Freishat (2014), for example, suggest that SET does not measure teaching effectiveness because a teacher's performance should be basically assessed in terms of the ability to facilitate learning, but measuring learning facilitation is hard, and the evaluations given by students after the course are only poor proxies. In addition, Emery et al. (2003) show that SET can fail to capture the teacher's ability to foster effective learning and is not conducive to the improvement of the educational outcome. Crumbley et al. (2001), in their examination of students' perception of the SET, show that poor SET scores may be a signal of inadequacy of student effort, of poor quality of the teacher's instructional input, or of both, and observe that SET scores may also be used by students as a retaliation against teachers for bad grades, heavy workloads, and so on. Meta-analyses of the literature even suggest that there might be no significant relationship between students' evaluations of a teacher's performance and the extent to which students actually learn from that teacher (Clayson 2009; Uttl et al. 2017). The main conclusion that can be drawn from this body of work seems to be that we are still lacking a clear conceptual framework and evaluation methodology to be able to assess to what extent SET is useful, for what specific purposes, and what are the essential limitations of its use.

In this paper, we move from the acknowledgement of the intrinsic limitations of SET as pointed out by the literature, but we also observe that, as SET is so widely used in the current educational practice, it is important to understand at least how to design incentive schemes for SET administration that bring about socially beneficial outcomes to some extent. This does not amount to legitimizing SET as the appropriate tool for the monitoring of instructional quality. Our goal is to contribute to the optimization of its use while waiting for a more solid scientific consensus for or against its adoption. In particular, we study 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 of the classroom. Moreover, 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 the framing of 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 in the evaluation process more than other types of students. Therefore, where educational systems work well, we expect higher levels of participation in, 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 its confidential character, and the fact that its 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). This establishes in turn 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 (that can in principle be overcome by using random sampling schemes for students, see Goos and Salomons, 2017), 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 tends to be sensitive to race and gender (Basow and Silberg 1987; Bavishi et al. 2010; Basow et al. 2013), and is 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. Fah and Osman (2011), analyzing the relationship between tutorial ratings, course and teacher characteristics in Canadian SET show that, whereas gender seems to have no effect on students' rating, ethnicity does. Students of Indian origin tend to assign higher course scores compared to other ethnicities, and this might depend on the fact that, as courses are taught in English, they encounter less problems than non-native English speakers from other overseas countries. Goos and Salomons (2017), moreover, point out that SET, and the related response rate, is affected by the disciplinary field and type of course attended by students, and that the response rate may be strongly improved when even a very small grade incentive is offered in exchange (see also, Dommeyer et al., 2004). A more ambiguous variable 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 social communication and influence factors (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). Finally, although the perceptions of students and teachers with regard to effective teaching are positively correlated, differences exist as well (Bosshardt and Watts 2001). For example, students care more about the teacher's preparation for class than instructors do. Pan et al. (2009) find that, contrary to widespread opinion, students tend to value the quality of teaching (e.g., ability to explain, facilitation of understanding) more than teachers' personality traits (e.g., sense of humor, charisma, extraversion, and so on).

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 modeling shortcoming in the attempt to understand under what conditions SET may be conducive to socially optimal results. We will now present a simple evolutionary game-theoretic model that provides a conceptual context for the study of the social selection of SET-driven optimal outcomes.

3 The model

Several game-theoretic models of teacher/student and teacher/teacher strategic interactions are currently available. Building on the seminal works 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 and of getting high rating. The model suggests that, since students are able to understand the quality of teachers, when professors are equal in every other respect, the teacher who offers more generous grading will tend to receive the highest student ratings. Alternatively, if the teachers are viewed as distinctively different by students, the one with the lower rating can offset the differential by easing up grading criteria, thus leading to grading inflation. Correa (2001) shows that this setting naturally leads to a social dilemma situation with the well-known sub-optimality issues. Correa (2003) considers 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 vs. less committed approach to teaching, and introducing the issue of diversity in players' capabilities and ethical standards. Strategic behavior of teachers is of 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.

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 overall evaluation is a weighted sum of the two. By providing low effort teaching output, teachers of the first course therefore make it less likely that the students are well prepared for the second course. Consequently, students obtain relatively worse evaluations for the second course, 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 any time t (time is continuous), a number of couples of teachers are matched to play an evaluation game, one for each matched couple. 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, on the basis of their performance in a subsequent, related course, retrospectively rate the first course professor's actual contribution to their preparedness for the second course. As far as 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 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 of teacher evaluation 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. Finally, we assume that the game is symmetric; that is, teachers do not differ in terms of skills, or other features that may generate asymmetries in payoffs. In terms of the evaluation game for the two teachers, the (symmetric) payoff matrix for stage *I* is:

$$\text{Stage } I : \begin{matrix} & D & E \\ D & \alpha & 0 \\ E & 1 & \beta \end{matrix} \tag{1}$$

where $1 > \alpha, \beta > 0$. 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 specifically a better one than the column teacher providing high effort (Krautmann and Sander 1999; Oleinik 2009). Accordingly, the worst outcome is the (D, E) one, where the teacher provides high effort and receives a worse evaluation than the other teacher providing low effort. Notice that 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) could be rationally preferred to (E, E) in that, in a long-term perspective, they have an interest in maximizing the knowledge return to their educational investment – a judgment which conflicts with their short-term interest to reward teachers who offer low effort courses, minimizing their study burden. In what follows, however, we restrict for simplicity the analysis to the case $\alpha < 1$.

At stage *II*, we assume that *D* 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:

$$\text{Stage } II : \begin{matrix} & D & E \\ D & a & 1 \\ E & 0 & b \end{matrix} \tag{2}$$

where $a > 0$ and $1 > b > 0$. 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. The relative size of parameters *a* and *b* depends on teachers’ attitudes toward their teaching duties, and ultimately on their motivations. More specifically, if $a < 1$, then the teacher choosing *D* gets a higher payoff if the other teacher provided a low effort teaching output *E*, with a consequent lower overall

preparation for the students. Vice-versa if $a > 1$. Analogously, if $a > b$ (this is always true in the context where $a > 1$), then both teachers obtain in (D, D) a higher payoff than in (E, E) ; vice-versa if $a < b$. Therefore, the nature of the strategic interaction is heavily affected by the teachers' normative orientation and thus, ultimately, by the 'cultural' factors that shape the prevailing set of effort-related social norms. As already anticipated in our discussion of the literature, this modeling feature points out how the analysis of the effectiveness of a SET scheme should always be related to the specific social environment in which it takes place. In different environments, teachers and students might be literally playing different kinds of games.

With respect to stage *I*, the payoff structure is now overturned, and in the single stage *II* 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. To compute the teachers' payoffs over the two stages, we assume that payoffs earned at stage *II* are weighted (discounted) by a factor $\theta \in (0, 1)$. Therefore, the overall payoffs are given by the matrix:

$$\text{Stage I \& Stage II: } \begin{array}{cc} & \begin{array}{c} D \\ E \end{array} \\ \begin{array}{c} D \\ E \end{array} & \begin{array}{cc} \alpha + \theta a & \theta \\ 1 & \beta + \theta b \end{array} \end{array} \quad (3)$$

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*. In the following section we will assume that the adoption process of strategies *D* and *E* is driven by a payoff-monotonic evolutionary dynamics, and we will highlight the dynamic regimes that may be observed under the payoff matrix (3).

4 Social selection dynamics

4.1 Evolutionary dynamics

Assume that the population of teachers is very large, and that at each time t (a number of couples, each one consisting of) two teachers are randomly selected 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 (courses need time to be prepared and syllabuses are published in advance). 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 $1 \geq x(t) \geq 0$. 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. Hence, according to the payoff matrix (3), the expected payoffs accruing to strategies *D* and *E* are given, respectively, by:

$$\pi_D(x) = [\alpha + \theta(a - 1)]x + \theta \quad (4)$$

$$\pi_E(x) = (1 - \beta - \theta b)x + \beta + \theta b \quad (5)$$

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):

$$\dot{x} = x(1 - x) [\pi_D(x) - \pi_E(x)] \tag{6}$$

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

$$\pi_D(x) - \pi_E(x) = \theta(1 - b) - \beta + [\alpha + \beta - 1 + \theta(a + b - 1)]x \tag{7}$$

The dynamic behavior of the replicator dynamics (6) 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 for every x 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 stationary state $\bar{x} \in (0, 1)$ of equation 6, where $\pi_D(\bar{x}) = \pi_E(\bar{x})$, corresponds to a mixed strategy Nash equilibrium of the static game (3), where both teachers play strategy D with probability \bar{x} , and strategy E with probability $1 - \bar{x}$; *ii*) the states $x = 0$ and $x = 1$ are locally attractive stationary states if, respectively, (D, D) and (E, E) are (strict) Nash equilibria of the two-stage game defined by the payoff matrix (3).

The social selection dynamics (6) describe 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 (6) can be classified as follows:

1. If $\pi_D(x) - \pi_E(x) \geq 0$ (respectively, $\pi_D(x) - \pi_E(x) \leq 0$) 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 of strategies $x(0) \in (0, 1)$, the trajectory starting from it 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 it approaches the attractive stationary state $x = 0$ (where all teachers play E).
2. If there exists a repulsive interior stationary state $\bar{x} \in (0, 1)$ (where both strategies coexist), separating the basins of attraction of the attractive stationary states $x = 0$ and $x = 1$, then we shall say that a *bistable dynamic regime* occurs.
3. If there exists an interior stationary state $\bar{x} \in (0, 1)$ and, for any initial distribution of strategies $x(0) \in (0, 1)$, the trajectory starting from it approaches \bar{x} , then we shall say that a *coexistence dynamic regime* occurs.

Note that the payoff differential (7) is a strictly increasing function of x if:

$$\alpha + \beta - 1 + \theta (a + b - 1) > 0 \quad (8)$$

If condition (8) holds, then the relative performance of strategy D , with respect to strategy E , improves as the share x of teachers adopting D increases, and vice-versa if (8) is strictly violated. The context in which (8) 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, the left-hand side of (8) measures how the social incentives at work tend to depend on the aggregate distribution of behaviors across teachers. In a situation where the payoff differential between D and E increases with the share x of teachers adopting D (and, accordingly, decreases with the share $1 - x$ of teachers adopting E), we have a ‘snowball’ social selection dynamics where the behavior that becomes socially prevailing eventually takes over at the expense of the other one. When on the contrary the payoff differential between D and E decreases with the share x of teachers adopting D , we have a ‘homeostatic’ social selection dynamics that tends to preserve diversity of behaviors across the population of teachers, and to reduce the relative share of a certain behavioral type if it becomes too prevalent. As already remarked, the social selection is entirely governed here by the frequency of adoption of the available behaviors, 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 the 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.

In order to gain more insight into the structure of the dynamic regimes of the model that will be presented below, it is convenient to have first a closer look into how the model’s parameters contribute to condition (8) being met or not. The validity of condition (8) (‘snowball’ social selection dynamics) is favored by relatively larger values of α and a . These are the parameters that control how rewarding it is for teachers to coordinate upon a difficult (D) course at stages I and II of the game, respectively. Likewise, the onset of the ‘snowball’ regime is also favored by relatively higher values of β and b , namely the rewards associated to coordinating upon an easy (E) course at stages I and II , respectively. The intuition behind this is clear: the more rewarding a given strategic option becomes, the more likely that the social dynamics will imply its widespread adoption once a critical mass of teachers has already adopted it. The structure of condition (8) also highlights the importance of the relative size of the rewards that characterize each stage of the game. The sign of the left-hand side of (8) depends in particular on whether the sum of the rewards from the coordinated outcomes at each stage, where both teachers play the same strategy (i.e. they both choose D or E) exceeds or not 1, namely, the sum of the payoffs from the non-coordinated outcomes where teachers choose different strategies at the same stage. Remember that in our payoff normalization, the latter quantity has been set constant to 1. In particular, the relative payoff from coordination vs. mis-coordination at stage II ($a + b - 1$) determines whether the discount factor θ has a positive or negative impact on the condition (8). When $a + b > 1$, that is, when the value of coordination

at stage *II* exceeds the value of mis-coordination, a higher discount factor favors the onset of the ‘snowball’ scenario where all teachers coordinate upon the same effort level, and where therefore the suboptimal equilibrium with low educational quality may eventually emerge. When $a + b > 1$ and most teachers choose low teaching effort, there is little incentive for the other teachers to go for high teaching effort, as unilateral deviations are relatively non-rewarding: hence the low effort equilibrium eventually prevails, and all the more so the higher the weight teachers place upon the payoffs from stage *II* (i.e., the higher θ). A similar reasoning makes the emergence of the high effort equilibrium likely when most teachers choose high effort under the same condition. On the contrary, if $a + b < 1$, the value of mis-coordination at stage *II* is relatively high, and this gives teachers an incentive to go for high effort when most teachers choose low effort or vice versa. Under this condition, the higher the weight teachers tend to assign to the payoffs at stage *II* (i.e., the higher θ), the more likely that they will go against the trend and choose a level of effort that is different from the one chosen by most teachers. As a consequence, a higher θ plays now against the onset of the ‘snowball’ dynamic regime and in favor of the ‘homeostatic’ one. In a nutshell, therefore, what condition says is that the ‘snowball’ regime will prevail whenever the total net value of coordination (that is, the value of coordination at stage *I* minus the value of mis-coordination at stage *I*, plus the weighted value of coordination at stage *II* minus the weighted value of mis-coordination at stage *II*) is positive. Vice versa, the ‘homeostatic’ regime will prevail when the total net value of coordination as defined above is negative (and therefore there is a stable incentive to mis-coordinate).

In the light of the above remarks, the formal characterization of the dynamic regimes as offered by Propositions 1 and 2 below is easily read and interpreted. In particular, Proposition 1 characterizes the ‘snowball’ regime and Proposition 2 the ‘homeostatic’ regime.

Proposition 1 *If condition (8) holds, then:*

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

$$\theta \geq \frac{\beta}{1 - b} \tag{9}$$

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

$$\theta \leq \frac{1 - \alpha}{a} \tag{10}$$

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

$$\frac{\beta}{1 - b} > \theta > \frac{1 - \alpha}{a} \tag{11}$$

The interpretation of the role of the parameters in the conditions of Proposition 1 is relatively straightforward. Consider for instance condition (9) for the dominance of *D*. The larger β and the closer b to 1 (i.e., the more rewarding ceteris paribus the coordination on the low effort strategy *E* at stages *I* and *II*, respectively), the less likely that dominance of *D* may occur, as the viable values of the weight θ

are more tightly constrained. Likewise, the larger α and a (i.e., the more rewarding ceteris paribus the coordination on the high effort strategy D at stages I and II , respectively), the less likely that dominance of E may occur, again due to tighter constraints on θ . The bistable pattern emerges when the two previous conditions for dominance are both simultaneously violated, and θ sits in an intermediate range of (feasible) values.

Proposition 2 *If condition (8) is strictly violated, then:*

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

$$\theta \geq \frac{1 - \alpha}{a} \tag{12}$$

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

$$\theta \leq \frac{\beta}{1 - b} \tag{13}$$

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

$$\frac{1 - \alpha}{a} > \theta > \frac{\beta}{1 - b} \tag{14}$$

The intuition for the interpretation of the conditions in Proposition 2 is an easy adaptation of that for the conditions in Proposition 1. Again, we have a condition for the dominance of D that is less likely met the smaller (ceteris paribus) the pay-off from high effort at both stages I and II , as this makes the constraint on the value of θ tighter; and a condition for the dominance of E that is less likely met the smaller (ceteris paribus) the payoff from low effort at stages I and II , for a similarly tightening constraint on θ . Coexistence occurs when both dominance conditions are simultaneously violated, and θ sits in an intermediate range of (feasible) values.

Proofs of Propositions 1 and 2 are straightforward. The various dynamic regimes are illustrated by Figs. 1, 2, 3, where full dots and empty dots represent, respectively,

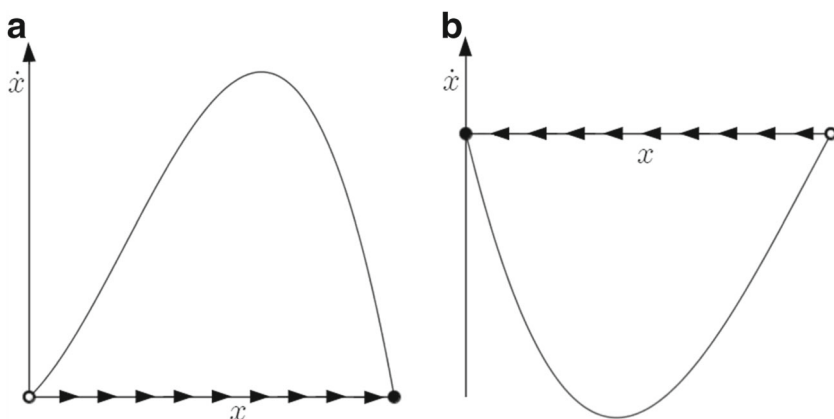


Fig. 1 Panel a: Takeover of high effort strategy D . Panel b: Takeover of low effort strategy

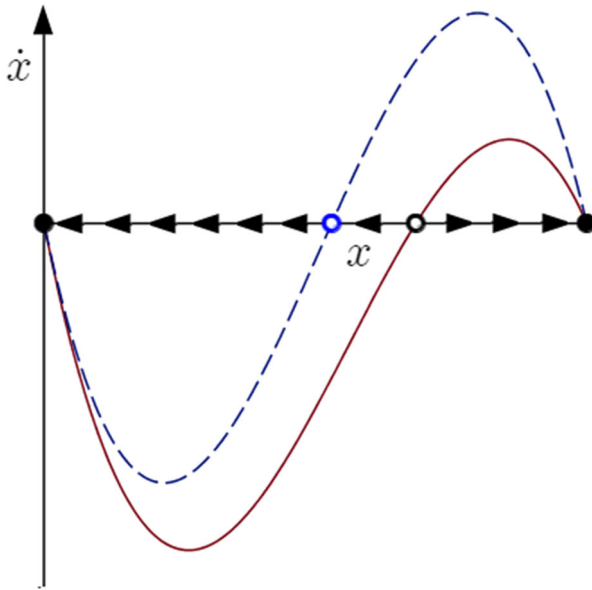


Fig. 2 Bistable regime. Arrows in figure are related to the red solid curve. The dashed blue curve shows the shift of the basin of attraction of the stationary state $x = 1$ induced by an increase in θ (notice the new \bar{x} depicted in blue)

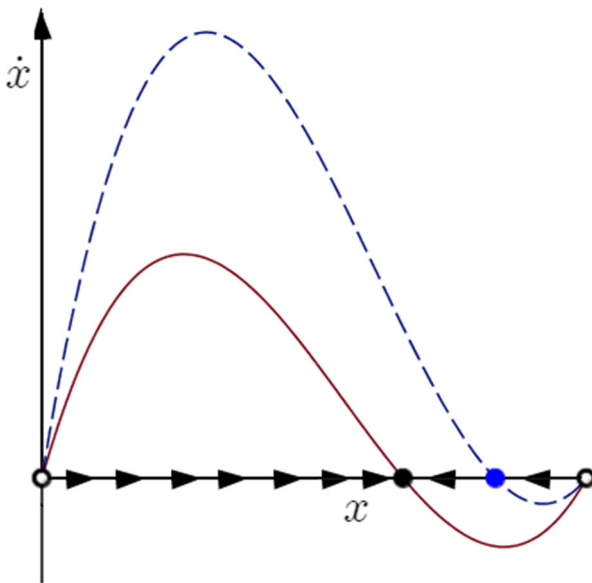


Fig. 3 Coexistence regime. Arrows in figure are related to the red solid curve. The dashed blue curve shows the increase in the share of teachers playing D induced by an increase in θ (notice the new \bar{x} depicted in blue)

attractive and repulsive stationary states. For cases (iii) of Propositions 1 and 2, the interior stationary state is given by:

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

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

Proposition 3 *It holds:*

$$\text{sign} \frac{\partial \bar{x}}{\partial \theta} = \text{sign} (1 + \alpha b - b - \alpha - \beta a) \quad (16)$$

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 equation 16, we have that:

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

The weight of stage II payoffs 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 social selection 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.

5 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) – that is, we prioritize their ‘rational’ long-term preferences over their possibly conflicting impulsive, short-term ones. 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. In view of the payoff structure (1)-(2), teachers’ payoffs evaluated at the stationary states $x = 0$ and $x = 1$ are respectively given by:

$$\begin{aligned} \pi_E(0) &= \beta + \theta b \\ \pi_D(1) &= \alpha + \theta a \\ \pi_D(\bar{x}) &= \pi_E(\bar{x}) = [\alpha + \theta(a - 1)]\bar{x} + \theta \end{aligned}$$

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:*

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

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:

$$\alpha + \theta(a - 1) > 0 \tag{18}$$

Notice that if $a \geq 1$, then (18) is always satisfied; if $a < 1$, then (18) holds if:

$$\theta < \frac{\alpha}{1 - a} \tag{19}$$

To understand the meaning of Proposition 4, let us consider the bistable dynamic regime, and remember that the difference $\alpha - \beta$ represents the payoff gain (or loss, if negative) that each teacher gets passing from the uniform high effort state (D, D) to the low effort state (E, E), in stage I of the game. Analogously, the difference $a - b$ represents the payoff gain (or loss, if negative) that each teacher gets passing from the uniform high effort state (D, D) to the low effort state (E, E), in stage II of the game. 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 (D, D), but due to the benefits of free riding the uniform low effort state (E, E) is the only Nash equilibrium of the stage I game. In this case, $\alpha - \beta$ is the welfare gain for each teacher from achieving the social optimum (D, D) instead of the Nash equilibrium (E, E), 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 $a > b$ and therefore they still prefer a uniform high effort state (D, D) to a low effort one (E, E) from the point of view of the students’ performance in the follow-up exam,¹ then condition (17) is trivially satisfied, and this means that the high effort equilibrium (D, D) is welfare improving upon the low effort one (E, E) in the bistable regime. In this case

¹That 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.

(which we could call the ‘goodwill scenario’), therefore, if the initial share of teachers choosing D is too small, the social dynamics eventually select the Pareto inferior equilibrium. If, on the contrary, $a < b$ holds, and therefore the 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 the uniformly low effort, a case that we could term the ‘direct responsibility scenario’), then the high effort equilibrium (D, D) will be Pareto optimal only if the stage II welfare gain $b - a$ from low effort is discounted enough by teachers, or if alternatively such gain is smaller than the stage I gain $\alpha - \beta$ 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 (E, E) from the perspective of both stages I and II (and thus, in particular, $\alpha < \beta$ and $a < b$ hold), for instance because the blame for the students’ performance in the follow-up exam is not put on the teachers according to the prevailing social norms (a case that we could term the ‘shirking scenario’), then condition (17) 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 peers (Kitts 2006). Finally, in the case where teachers prefer a uniform low effort state (E, E) from the point of view of stage I (i.e., $\alpha < \beta$) but prefer a uniform high effort state (D, D) from the point of view of stage II (i.e. $a > b$, e.g. because despite their weak 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 $\alpha - \beta$ from a uniform high effort state at stage I and the (discounted) welfare gain $\theta(b - a)$ 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 state 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 (that is, if the initial distribution of strategies is $x(0) = 1$).

According to condition (18), the higher the payoffs from the uniform high effort state (D, D) at stages I and II , respectively α and a , the more likely that (18) is met. Notice that condition (18) is always satisfied if $a \geq 1$, that is, if in the stage II game we have that playing D against D gives a higher payoff than playing D against E . If $a < 1$, then the smaller θ (i.e. the more the payoff at stage II is discounted) the

more likely that the condition (18) is met. If teachers' payoffs associated to (D, D) are high enough, in both stages of the game, then they prefer the high effort stationary state $x = 1$ to the interior stationary state \bar{x} , where both strategies coexist, 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. The opposite holds if α and a are low enough, that is, if teachers are not motivated about the overall performance of their students at both stages *I* and *II*.

Notice moreover how conditions (17)-(18) are compatible with conditions (11)-(14), which identify the bistable and coexistence dynamic regimes, respectively, but are not implied by them. This amounts to say 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.

6 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 shows how the effectiveness of SET in fostering the emergence of high effort equilibria (that best serve the long-term interest of students) is significantly improved when teachers have an intrinsic motivation to provide high effort. Attaching an intrinsic value to the offering of quality education reduces teachers' benefit to free ride by providing a low effort course, and reinforces the incentive related to teachers' accountability for the students' performance in the follow-up exam. 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 teachers 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 alternatively one specific effort level 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 specifically 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 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.

It should also be remarked that SET need not be the only possible way to incentivize teachers to provide high effort courses. The experiment undertaken in countries such as Finland in terms of de-structuring educational programs to allow teachers and students to build their own approach in an open-ended, self-responsible way, is particularly interesting in this regard (Sahlberg 2015), also in view of its considerable success in helping Finnish students to achieve very high PISA scores. It must be noted, however, that such an achievement is also made possible by the high social prestige of the teacher role (Hargreaves 2009) and by the egalitarian orientation of Finnish society in terms of equality of opportunity (Sahlberg 2012) – the combined effect of which certainly motivates Finnish teachers to offer quality courses and to expect commitment to learning by students without the need to rely upon formal accountability systems like SET (Toom and Husu 2012). This is an example of how, in certain socio-cognitive contexts, excellence in education may be successfully pursued through mainly intrinsic rather than extrinsic incentives. The relative effectiveness of formal accountability schemes like SET vs. alternative forms of social control in educational systems within different cultural contexts is a still under-researched topic that would deserve more attention.

Placing our results in the context of the existing literature yields some interesting implications. First of all, the analysis of social incentives shows clearly how the evaluation of the effectiveness of a given SET scheme must necessarily be content specific, and that the same scheme may lead to different levels of effectiveness in promoting educational quality depending on the environmental conditions. Secondly, in certain circumstances even small changes in the structure of incentives may bring

about substantial changes in the dynamic regime (and in the long-term outcomes), whereas in other cases even relatively large changes do not essentially alter the social dynamics — it all depends on how close the system parameters are to the frontier that separates different regimes (or their sub-regimes as described by Propositions 1 and 2). Finally, the welfare implications of changes in the structure of incentives may be very complex, and without a proper understanding of the underlying structure of the strategic interaction policy interventions may not provide the desired results. The basic message behind our model is therefore that keeping social incentives explicitly into account may substantially alter our analysis and assessment of SET schemes in specific socio-cultural contexts and under specific circumstances.

Our model presents the simplest possible version of a social selection dynamics of teachers' choices, but clearly one can also consider 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 the 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 explores this promising path in its full potential.

Funding Open access funding provided by degli Studi di Sassari within the CRUI-CARE Agreement. The research of Angelo Antoci was supported by the Italian Ministry of Education (MIUR) ("Dipartimenti di Eccellenza" Program 2018-2022).

Compliance with Ethical Standards

Conflict of interests The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

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. *Coll Stud J* 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. *J Divers High Educ* 3:245–256
- Becker WE, Watts M (1999) How departments of economics evaluate teaching. *Am Econ Rev* 89:344–349
- Boring A (2017) Gender biases in student evaluations of teaching. *J Public Econ* 145:27–41
- Bosshardt W, Watts M (2001) Comparing student and instructor evaluations of teaching. *J Econ Educ* 32:3–17
- Boswell SS (2016) Ratemyprofessors is hogwash (but I care): Effects of Ratemyprofessors and university-administered teaching evaluations on professors. *Comput Hum Behav* 56:155–162
- Braga M, Paccagnella M, Pellizzari M (2014) Evaluating students' evaluations of professors. *Econ Educ Rev* 41:71–88
- Brodie DA (1998) Do students report that easy professors are excellent teachers? *Canadian Journal of Higher Education* 28:1–20
- Brown BW, Saks DH (1987) The microeconomics of the allocation of teachers' time and student learning. *Econ Educ Rev* 6:319–332
- Cahuc P, Carcillo P, Zylberberg A (2014) *Labor economics*. MIT Press, Cambridge
- Carrell SE, West JE (2010) Does professor quality matter? Evidence from random assignment of students to professors. *J Polit Econ* 118:409–432
- Catsiapis G (1987) A model of educational investment decisions. *Rev Econ Stat* 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. *Econ Educ Rev* 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. *J Mark Educ* 31:16–30
- Correa H, Gruver GW (1987) Teacher-student interaction: A game theoretic extension of the economic theory of education. *Math Soc Sci* 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
- Crumbley L, Henry BK, Kratchman SH (2001) Students' perceptions of the evaluation of college teaching. *Qual Assur Educ* 9(4):197–207
- Darwin S (2017) What contemporary work are student ratings actually doing in higher education? *Studies in Educational Evaluation* 54:13–21
- Dommeyer CJ, Baum P, Hanna RW, Chapman KS (2004) Gathering faculty teaching evaluations by in-class and online surveys: Their effects on response rates and evaluations. *Assessment and Evaluation in Higher Education* 29(5):611–623
- Emery C, Kramer T, Tian R (2003) Return to academic standards: A critique of student evaluations of teaching effectiveness. *Qual Assur Educ* 11(1):37–46
- Entwistle NJ, Thompson J, Wilson JD (1974) Motivation and study habits. *High Educ* 3:379–396
- Entwistle N, Hanley M, Hounsell D (1979) Identifying distinctive approaches to studying. *High Educ* 8:365–380
- Ewing AM (2012) Estimating the impact of relative expected grade on student evaluation of teachers. *Econ Educ Rev* 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
- Fah BCY, Osman S (2011) A case study of student evaluation of teaching in university. *International Education Studies* 4(1):44–50
- Figlio DN, Kenny NW (2007) Individual teacher incentives and student performance. *J Public Econ* 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
- Goos M, Salomons A (2017) Measuring teaching quality in higher education: assessing selection bias in course evaluations. *Res High Educ* 58:341–364
- Gotlieb J (2013) Variables that can affect student ratings of their professors. *Journal of Learning in Higher Education* 9:19–27
- Greenwood GE, Ramagli H. J. Jr (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
- Hargreaves L (2009) The status and prestige of teachers and teaching. In: Saha L. J., Dworkin A. G., Hargreaves L (eds) *International handbook of research on teachers and teaching*. Springer Science, New York, pp 217–229
- Hartman KB, Hunt JB (2013) What ratemyprofessors.com reveals about how and why students evaluate their professors: A glimpse into the student mind-set. *Mark Educ Rev* 23:151–161
- Hernández-Julián R, Looney A (2016) Measuring inflation in grades: An application of price indexing to undergraduate grades. *Econ Educ Rev* 55:220–232
- Hofman E (2014) An obligation of conscience: Gossip as social control in an Eighteenth-century Flemish town. *Eur Rev Hist* 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. *Econ Educ Rev* 37:148–164
- Kherfi S (2011) Whose opinion is it anyway? Determinants of participation in student evaluation of teaching. *J Econ Educ* 42:19–30
- Kitts JA (2006) Collective action, rival incentives, and the emergence of antisocial norms. *Am Sociol Rev* 71:235–259
- Knol M.H., 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. *Comput Hum Behav* 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. *Soc Psychol Educ* 15:337–351
- Krautmann AC, Sander W (1999) Grades and students' evaluation of teachers. *Econ Educ Rev* 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. *Econ Educ Rev* 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. *Comput Hum Behav* 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. *East Econ J* 38:296–309
- McCready DJ (1981) Student evaluation of teaching. *Can J High Educ* 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. *J Econ Behav Organ* 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
- Pan D, Tan GSH, Ragupathi K, Booluck K, Roop R, Ip YK (2009) Profiling teacher/teaching using descriptors derived from qualitative feedback: Formative and summative applications. *Res High Educ* 50:73–100
- 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. *J Occup 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. *Rev Educ Res* 49:497–511
- Rynes SL, Gerhard B, Parks L (2005) Personnel psychology: Performance evaluation and pay for performance. *Annu Rev Psychol* 56:571–600
- Sahlberg P (2012) A model lesson: Finland shows us what equal opportunity looks like. *American Educator* 36:20–27
- Sahlberg P (2015) *Finnish lessons 2.0. What can the world learn from educational change in Finland?* Teachers College Press, New York
- 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. *J Educ Comput Res* 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
- Stark PB, Freishtat RL (2014) An evaluation of course evaluations. *Science open research*. <https://doi.org/10.14293/S2199-1006.1.-AOFRQA.v1>.
- Sun Y (1998) The academic success of East-Asian-American students: An investment model. *Soc Sci Res* 27:432–456
- Toom A, Husu J (2012) Finnish teachers as 'makers of the many'. Balancing between broad pedagogical freedom and responsibility. In: Niemi H, Toom A, Kallioniemi A (eds) *Miracle of education: The principles and practices of teaching and learning in Finnish schools* (pp 41-56). Sense Publishers, Rotterdam
- Uttl B, White CA, Wong Gonzalez D (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. *Econ Educ Rev* 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
- Wilson R (1998) New research casts doubt on value of student evaluation of professors. *Chronicle of Higher Education* 44(19):A12–A14
- 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