Working Paper No. 27

Optimal Grading

Robertas Zubrickas

Universität Zürich
ISU – Institut für Strategie und Unternehmensökonomik

UNIVERSITÄT BERN

Leading House Working Paper No. 27

Optimal Grading

Robertas Zubrickas

March 2008

The Swiss Leading House on Economics of Education, Firm Behavior and Training Policies is a Research Programme of the Swiss Federal Office for Professional Education and Technology (OPET).

www.economics-of-education.ch

Optimal Grading

Robertas Zubrickas*
Stockholm School of Economics

March 2008

Abstract

In the framework of static mechanism design games with non-pecuniary rewards, we solve for optimal student grading schemes and attempt to explain the observed mismatch between students' grades and their abilities. The model predicts that the more pessimistic the teacher is about her students, the more generous she should be in grading them. Generally, the "no distortion at the top" property ceases to hold for optimal contracts with costless non-pecuniary rewards, and we argue that the compression of ratings as witnessed in job performance appraisals could be an equilibrium outcome. The presented theoretical findings are strongly supported by empirical evidence from the related literature in psychological and educational measurement.

Keywords: Mechanism design; non-pecuniary incentives; optimal grading schemes; mismatch of grades and abilities; compression of ratings.

JEL-codes: D82, D86, I20, J41.

1 Introduction

The vast literature on subjective evaluation has long dealt with the phenomenon of the compression of ratings, which is about supervisors' shallow differentiation of good from bad performance of their subjects by

Acknowledgement: I am indebted to Tore Ellingsen and Magnus Johannesson for their guidance into this research. This paper has also benefited from the discussions with Drew Fudenberg, Jean Tirole, Jörgen Weibull, and, especially, Karl Wärneryd. I thank the European Network for Advancement of Behavioural Economics, and the Jan Wallander and Tom Hedelius Foundation for financial support. *Contact*: Department of Economics, Stockholm School of Economics, Box 6501, SE-11383 Stockholm, Sweden. E-mail: robertas.zubrickas@hhs.se.

means of ratings (see, *e.g.*, Prendergast, 1999, for an economist account on the issue, or Murphy and Cleveland, 1995, for a review of related studies from the psychological strand of the literature). Within this literature, we can also position the phenomenon of mismatch or low correlation between students' university grades and their actual abilities, which similarly raises the question why teachers turn out to be too generous in grading their students (Goldman and Widawski, 1976, and Johnson, 2003). In general, and quite surprisingly, these phenomena universally persist in different settings despite that they seemingly lead to inefficient outcomes of principal-agent relationships. Evidently, given a uniform rating or grading scheme, agents would put just enough effort to reach some minimum performance standard granting the desired reward, but, simultaneously, there must be some rationale behind those enduring coarse ranking schemes. However, neither the psychological strand of the literature nor that of economics provides any rigorous explanation of these phenomena, where the argument is typically somewhat heuristic (see Prendergast, 1999, Murphy and Cleveland, 1995, or Johnson, 2003).

In this paper, we attempt to tackle the raised phenomena through the perspective of a static principal-agent model with hidden information featuring costless transfers between the parties. In particular, we treat rating or grading standards as an implicit contract between the principal and the agent, the crucial feature of which is a costless non-pecuniary reward from the principal to the agent in return for the agent's exerted (observable and verifiable) effort to the principal's advantage. To emphasize, the reward enters only the payoff function of the agent, and the principal herself bears no cost (at least, no variable cost) of compensating the agent. A more thorough discussion why this assumption of the model is realistic is postponed till later in the text, and here we only mention that the situations applicable to the model would feature a non-pecuniary reward such as praise or, more tangibly, a grade at school or a rating in a job performance appraisal, which can motivate the agent to put more effort.

For ease of exposition, from the very outset we build the model around the premises of a concrete example about a teacher's designing grading rules to evaluate the performance of her students, even though we generally aim at a broader class of related agency problems (because of that the pairs of terms "teacher-student" and "principal-agent" are used synonymously in the paper). Accordingly, the purpose of this paper is to design optimal grading schemes on the part of the teacher who aims at extracting the highest expected effort level from her grade-minded students. Based on the obtained theoretical findings, we shall argue that the observed mismatch between students' grades and their abilities, and

the compression of ratings, more generally, could be the optimal solutions to particular agency problems. Importantly, as shown later, the existing empirical evidence supports the conclusions of the model.

Despite being theoretically oriented in the conventional economics sense, this paper has, however, been largely motivated by findings from behavioral and experimental economics literature on contractual relationships. There is a growing number of theoretical and empirical studies arguing that along with pecuniary incentives people also tend to care about non-pecuniary motives and rewards, associated with the contract's design and implementation, such as the resulting self-esteem or importance of the agent, his perceived trustworthiness, or the principal's elicited praise and esteem (Brennan and Pettit, 2004, Frey and Osterloh, 2002, Berg *et al*, 1995, and Falk and Kosfeld, 2006). As a result of that, the standard principal-agent model with merely monetary incentives may render unfit to account for variation in effort levels exerted by, say, employees on fixed-wage jobs with no other present or future pecuniary stimuli, as already discussed in Akerlof (1982). The obvious direction for further research has become to enrich the standard model with more elaborate mechanisms aimed at capturing more closely the complexities of a principal-agent relationship (*e.g.*, Sliwka, 2007, and Benabou and Tirole, 2003). However, when thinking of a principal-agent model with non-pecuniary rewards, the most natural framework and application to bring forward is a teacher-student relationship, which this paper actually does; though, we shall also argue that the model, presented here, can be applied to other similar frameworks as well.

The proposed refinement that, unlike the agent, the principal is indifferent to the transfer between them is by no means new in the contract theory literature. It was formally studied, for example, in Guesnerie and Laffont (1984), one of the founding articles on mechanism design aimed at providing an all-encompassing solution to a broadly defined principal-agent problem. In particular, they distinguish between "type A" and "type B" preferences of the principal, where with the former preferences the principal's utility does not depend on the transfer between the parties, while the latter preferences are those of a more conventional appearance with costly transfers. In their study, however, the "type A" preferences are mainly discussed with the social planner's problem to solve in mind, in which the transfer between the parties is equivalent, literally speaking, to putting money from one pocket into another of the same jacket leaving the social welfare intact. As it will be argued, the solution method, as suggested in Guesnerie and Laffont (1984) to solve the principal-agent problem with the "A type" preferences, does not apply to our case, where the principal is more, in fact, of "type B", *i.e.*,

caring only about her own well-being with a fortunate feature that she does not pay for motivating the agent.

Nor does this paper stand alone in designing optimal grading rules.[1] Dubey and Geanakoplos (2005) also target the same problem but from a different perspective: they model a teacher-student relationship as a game of status with private information and stochastic output similarly as in a tournament. Hence, unlike in our model, they start with a multiple-agent setting, where an agent's utility from a grade depends on his or her class ranking, *i.e.*, status, the rewarded grade results in, but not on the grade *per se*. Another difference between the two models is that in Dubey and Geanakoplos (2005) the teacher designs grading rules in order to induce all her students to put in the maximal effort, which produces a stochastic output, while in our model it is the highest expected effort level that the teacher attempts to maximize, and there is no stochastic component in output. Not surprisingly, given these modeling differences, we draw different conclusions about optimal grading schemes. Dubey and Geanakoplos (2005) find that teachers should use coarse grading schemes and "pyramid" the allocation of grades: in equilibrium the highest grade would be available to fewer students than the second-highest grade, and so on. Our model similarly predicts that teachers should apply coarse grading schemes, but the level of "coarsening" depends on the distribution of student abilities. Furthermore, we don't find "pyramiding" to be an optimal grade allocation, especially, when there is a large mass of less able students in the class.

According to our theoretical findings, it could be optimal for teachers to appear excessively generous with their students, which, in fact, has been a big concern in the literature of educational measurement. We show that with an intention to maximize the average effort of her students the teacher who is pessimistic about their abilities should be more generous when grading them than she would have been if she had held higher expectations about their abilities. All this can lead to uneven distributions of grades among classes that are different in students' abilities, and to a mismatch and low correlation between grades and students' actual abilities. In general, extrapolating this paper's theoretical findings to other settings featuring non-pecuniary rewards such as in job performance appraisals with ratings, we shall argue that the compres-

[1] Though, it needs to be reckoned that not much theoretical work has been done on modeling a teacher-student relationship as a principal-agent model on its own, whereas this relationship has typically been modeled as a part of a more global game involving potential employers or university administration (see, *e.g.*, Ostrovsky and Schwarz, 2005). At the same time, more research has been done on the empirical side of the problem (see, *e.g.*, Johnson, 2003).

sion of ratings may turn out to be an optimal solution for the principal whenever she is uncertain about each agent's abilities or is not allowed to discriminate among her agents by offering ability-specific contracts.

As for empirical evidence, Goldman and Widawski (1976) report a negative correlation between students' Scholastic Aptitude Test scores (could be seen as a proxy measure of student abilities) and the grading standards at the classes the students were majoring in. According to this study (conducted at University of California, Riverside), the observed negative correlation is due to the fact that professors in a field containing students with high abilities tend to grade more stringently than do professors in a field with lower-ability students—precisely as our model predicts. These empirical findings were confirmed by similar studies conducted at Dartmouth College (Strenta and Elliott, 1987) and at Duke University (Johnson, 2003).

The paper is organized as follows. Section 2 introduces the model, which is formally solved in Section 3. Section 4 discusses the main findings and relates them to the phenomena raised in the introduction. Section 5 reviews the existing empirical evidence, and Section 6 concludes.

2 The Model

This section presents a principal-agent model with non-pecuniary rewards, which, for expositional tractability, closely follows, when possible, the textbook variants of related static models with adverse selection and the single agent as in, *e.g.*, Bolton and Dewatripont (2004) or Fudenberg and Tirole (1991).

Consider a teacher who has to set up grading rules for the students enrolled in her class.[2] A grading rule assigns grades—that are costless for the teacher to reward—to students' performance levels, which are verifiable and assumed to be perfectly correlated with their costly learning efforts. The teacher believes that it is their own grades that the students care about, and that they incur only disutility from studying.[3] In contrast to students' wants, the teacher wants all her students to

[2] To get round the discussion why students select particular courses, we can assume that our attention is only on compulsory courses, and students' choice of a major in college is based on considerations other than the expected grade average. This assumption is in line with empirical evidence that ability sorting among majors primarily takes place because of students' intrinsic preferences for particular majors (and not because of expected future earnings, which are, on the other hand, correlated with university grades), see, *e.g.*, Arcidiacono (2004).

[3] By this, we assume that the expected distribution of grades in the class does not affect a student's utility from a targeted grade and, accordingly, his or her learning effort choice decision. This assumption allows us to treat our model as a single-agent model. An extension to a multiple-agent problem is left for future research.

put as much effort as possible for a given grade. However, when setting grade-for-performance (equivalently, grade-for-effort) schemes, the teacher also needs to balance the incentives made available to her students, who may have different learning abilities but are all identical in other respects, including their attempts to save on effort needed for a given grade. The teacher knows the distribution of students' abilities, say, from her previous experience, but she cannot tell what abilities a particular student has.[4] Hence, assuming that she equally cares about every student's effort, the teacher's goal becomes to maximize the average effort in her class subject to the customary individual rationality and incentive compatibility constraints as more succinctly described below.

Using the direct mechanism approach with truthful revelation, the teacher designs a set of effort-grade allocations $\{x(\theta), r(\theta)\}_{\theta\in[\theta_a,\theta_b]}$, where θ is a student's ability measure distributed according to a twice differential cumulative distribution function F over $[\theta_a, \theta_b]$ with the probability density function f; the grade function r maps the ability set $[\theta_a, \theta_b]$ into $[0, 1]$ so that the maximum grade the teacher can offer is 1[5]; and the effort function x maps $[\theta_a, \theta_b]$ into a bounded interval $[0, \bar{x}]$, which, when needed, is assumed to be large enough to allow for an interior solution. Upon observing the set of available effort-grade allocations, a θ-type student optimally selects a type $\hat{\theta}$ to report to the teacher, who, subsequently, demands to put the effort $x(\hat{\theta})$ in return for the grade $r(\hat{\theta})$.

The implemented allocation $(x(\hat{\theta}), r(\hat{\theta}))$ results in the teacher's utility of

$$U_P(x(\hat{\theta}), r(\hat{\theta})) \equiv V(x(\hat{\theta})),$$

which is increasing in the effort x. Accordingly, the θ-type student enjoys

[4]Here, we impose a hidden information structure in our model. However, if it seems restrictive, as it could be thought of when considering teacher-student relationships, then we can, alternatively, allow that the principal is able to tell a student's type (similarly to Benabou and Tirole, 2003), but at the same time we would require that the teacher cannot discriminate among her students by applying ability-specific grading rules. With this alternative formulation, the optimization problem remains intact as in the case with the adverse selection framework, which is, therefore, retained for its link with the existing literature.

[5]Without an upper bound on the costless reward function, the optimal grading rule would be to demand the maximal feasible effort from every type of students and to reward them whatever abundantly. Putting an upper bound on the reward function not only remedies the arisen implausibility of the solution, but it is also a most natural thing to impose when considering teacher-student relationships, where there is typically a formal, institutionally set highest grade. Similarly, job performance is also normally appraised on a finite rating scale, and, finally, even praise, which could be thought as formally unbounded, may still have only a limited effect on the agent's utility resulting from it.

the utility of

$$U_A(x(\hat{\theta}), r(\hat{\theta}), \theta) = r(\hat{\theta}) - C(x(\hat{\theta}), \theta),$$

which needs to be at least as large as his reservation utility (henceforth, normalized to 0 for all types of students), and where $C(x, \theta)$ is the cost function measuring disutility from putting an effort x, with the properties $C_x > 0$, $C_{xx} > 0$, and $C_{x\theta} \leq 0$ (*i.e.*, the marginal cost from effort is lower for more talented students).

Finally, assuming that the principal and student are both rational own-utility maximizers, the optimization problem of the principal is to find the set of allocations $\{x(\theta), r(\theta)\}$ such that for every θ and $\hat{\theta}$ it maximizes

[Program 1]

$$\int_{\theta_a}^{\theta_b} V(x(\tilde{\theta})) f(\tilde{\theta}) d\tilde{\theta}$$

subject to

$$r(\theta) - C(x(\theta), \theta) \geq r(\hat{\theta}) - C(x(\hat{\theta}), \theta), \qquad \text{(IC)}$$

$$r(\theta) - C(x(\theta), \theta) \geq 0, \qquad \text{(IR)}$$

$$0 \leq r(\theta) \leq 1,$$

where (IC) stands for the incentive-compatibility constraint, which makes sure that it is optimal for the agent to report truthfully, *i.e.*, $\hat{\theta} = \theta$ in equilibrium, and (IR) is the individual-rationality or participation constraint, and the last constraint imposes an upper bound on the reward function.

To justify the main feature of the model that it costs nothing for the teacher to grade her students, or, at least, the cost is the same irrespective of the grade rewarded, but at the same time grades are of a value to students, we need to tackle two questions: (1) if there are no hidden costs of pecuniary nature; and (2) why students actually care about grades. With regard to the first question, one could think of reputational concerns that teachers may face while setting up grading standards. Students aiming at higher grades may base their decision whether to enroll in a particular course on the stringency of the applied grading standards simultaneously putting pressure on teachers to soften their grading rules and lure more students to their classes (see, *e.g.*, Johnson, 2003). However, as already discussed in footnote 2, students seem to select a major based on their personal preferences for it rather than on his or her expected grade at the selected major. Therefore, at least at major courses, teachers' reputational concerns should be of a lesser magnitude, hence, with the focus on major courses, as can be assumed henceforth, should

alleviate this problem. Furthermore, neither will the empirical evidence of the model be prejudiced, for the presented evidence mainly comes from the grading standards used at major courses. As for the "demand" side, *i.e.*, the second question, we can safely argue that students are likely to derive higher utility from a higher grade everything else equal due to, for instance, better signals sent to their parents, friends or even to themselves about their personal characteristics.[6]

Prior to solving the program, we place some further structure on the model by commonly assuming (A1)-(A3) that are loosely defined below. First, (A1), the single-crossing property holds (in fact, it has been imposed by requiring $C_{x\theta} \leq 0$). Second, (A2), we impose the assumption of the monotone increasing hazard rate, defined as $h(\theta) = f(\theta)/(1 - F(\theta))$, and the assumption implies that $h'(\theta) > 0$. Finally, (A3), without going into much detail, we require the functional forms of V and C be such that, when needed, the second-order conditions are fulfilled, in particular, we require that $C_{xx\theta} \leq 0$.

3 Solution

The standard adverse selection principal-agent model with monetary (*i.e.*, costly) transfers is solved using the method due to Mirrlees (1971), the main idea of which is to obtain a functional equation with one unknown by merging the agent's optimization problem with that of the principal. However, in our case with costless non-pecuniary rewards this method needs to be re-examined because the intercomparison of the parties' utilities is no longer possible, for, formally speaking, there is no linking term between the two utility functions, as is the transfer function in the standard model.[7] Instead, we shall approach Program 1 through its discrete version, as suggested in the next subsection, and, then, we shall take the limit of the obtained discrete-case results to arrive at the

[6] Ideally, one would want to think of grades as ability signals to the labor market, but then our model would need to be closed by introducing one more stage, the recruitment stage (*e.g.*, Ostrovsky and Schwarz, 2005). However, this extension would complicate things a lot, because it would eventually require elaborating on the school entry decision, too. Therefore, we ignore this discussion by just saying that once at school students tend to care about grades.

[7] Arguably, we could express our optimization problem in the terms of the standard framework by defining the principal utility function as $U_P(x, r) = V(x) - \lambda r$, and solve the problem for an arbitrary non-zero λ, and then to obtain the solution to Program 1 by taking the limit $\lambda \to 0$. However, we would then also need to modify the solution method due to an upper bound on the reward function, and to offer an elaborate discussion about any possible discontinuities at the limit. The main reason for proposing a different way of solving the program is the idea to approach the problem directly without firstly reverting to a costly-transfer case. Arguably, the suggested method also provides a clearer intuitive account of the obtained solution.

general solution with the continuum of agent types.[8] In addition, to make the problem more mathematically tractable we assume the following functional forms:

$$V(x) = x \qquad \text{and} \qquad C(x, \theta) = \frac{g(x)}{\theta}. \tag{1}$$

As will be argued later, the main properties of the obtained solution are by no means influenced by the assumed functional forms of the teacher's utility function V and student's effort cost function C. As for the effort cost function, we basically assume that the function C is separable in the effort x and the type θ, *i.e.*, $C = g(x)t(\theta)$, but once this granted, the further simplification $g(x)/\theta$ is just a matter of convenience. As for the teacher's linear utility in the student's effort, it implies that the teacher attaches equal weights to marginal effort increases of all her students, *i.e.*, she cares equally about every student. Later in the text, this case is considered as a benchmark against two other cases when V is strictly concave—the teacher values more the marginal effort increases of low performers than those of high performers—and when V is strictly convex (but not too convex)—the teacher, on the contrary to the previous case, more values increases in the performance of more able students.[9]

Discretization

As suggested previously, we discretize the student type space $[\theta_a, \theta_b]$ into n equal-length intervals:

$[\theta_a, \theta_a + \partial\theta], ..., [\theta_a + i\partial\theta, \theta_a + (i+1)\partial\theta], ..., [\theta_a + (n-1)\partial\theta, \theta_b]$, where $\partial\theta = (\theta_b - \theta_a)/n$. Accordingly, we discretize the initial (continuous) distribution F for students' abilities by setting $p(\theta_i) = \int_{\theta_a + (i-1)\partial\theta}^{\theta_a + i\partial\theta} f(\theta) d\theta$, $i = 1, ..., n$, which is the probability mass of student types within the interval $[\theta_a + (i-1)\partial\theta, \theta_a + i\partial\theta]$. In particular, in the probability $p(\theta_i)$ we understand by θ_i the starting point of the interval $[\theta_a + (i-1)\partial\theta, \theta_a + i\partial\theta]$,

[8] In footnote 4, we mentioned that we could treat the adverse selection framework as the approximation of the perfect information framework with the condition that all the students irrespective of their abilities are subject to the same grading rules. Hence, with perfect information, if there are a finite number of students in the class, then we exactly face a discrete-type problem, as defined below, and the limiting continuous case is just a convenient way of summarizing the properties of the solution to the discrete-type problem.

[9] The discussion about the functional form of the teacher's utility function is relevant in the light of the recent merit-pay programs aimed at improving teachers' incentives (see Lavy, 2002, Atkinson *et al*, 2004, or Lazear, 2003). Arguably, it is in the hands of the social planner to affect the functional form of the teacher's utility function when designing pay-for-student-performance schemes.

i.e., $\theta_i = \theta_a + (i-1)\partial\theta$, $i = 1, ..., n$. This kind of discretization will allow us to switch to the continuous case by taking the limit $n \to \infty$.

Correspondingly, the discrete version of Program 1 is defined below, where the teacher is to offer the contract $\{x(\theta_i), r(\theta_i)\}_{i=1}^n$ that maximizes for $i = 1, ..., n$

[Program 2]

$$\sum_{i=1}^{n} p(\theta_i)x(\theta_i)$$

subject to

$$r(\theta_i) - C(x(\theta_i), \theta_i) \geq 0, \qquad (IR_i)$$

$$r(\theta_i) - C(x(\theta_i), \theta_i) \geq r(\theta_j) - C(x(\theta_j), \theta_i), \qquad j \neq i, \qquad (IC_i)$$

$$0 \leq r(\theta_i) \leq 1.$$

First-order conditions

As it is standard in the principal-agent problems with adverse selection, the only individual rationality constraint that binds is IR_1, *i.e.*, that of the least efficient agent. After imposing that the effort $x(\theta_i)$ is non-decreasing in the student's type, which will have to be checked separately, due to the Spence-Mirrlees property we can restrict our attention to the following set of downward binding adjacent IC constraints:

$$r(\theta_i) - C(x(\theta_i), \theta_i) = r(\theta_{i-1}) - C(x(\theta_{i-1}), \theta_i), \qquad i = 2, ..., n. \qquad (2)$$

Observing that it must be optimal for the teacher to reward at least someone with the highest grade of 1 since it costs nothing to her, hence, in equilibrium the reward to the most efficient type has to be 1, *i.e.*, $r(\theta_n) = 1$. Next, supposing that the solution takes the form of a separating equilibrium, we can combine all the binding constraints (2) together with $r(\theta_n) = 1$ into one constraint, by doing which we eliminate the reward function r from the program rendering the following implementability constraint:

$$1 - \sum_{i=1}^{n} C(x(\theta_i), \theta_i) + \sum_{i=2}^{n} C(x(\theta_{i-1}), \theta_i) = 0. \qquad (3)$$

The Lagrangian of the reduced optimization program (without the monotonicity constraint on the effort x) is defined as

$$L(\{x(\theta_i)\}_{i=1}^n, \mu) = \sum_{i=1}^{n} p(\theta_i)x(\theta_i) + \mu(1 - \sum_{i=1}^{n} C(x(\theta_i), \theta_i) + \sum_{i=2}^{n} C(x(\theta_{i-1}), \theta_i)).$$

The first-order conditions with respect to effort levels $x(\theta_i)$ are

$$p(\theta_i) = \mu(C_x(x(\theta_i), \theta_i) - C_x(x(\theta_i), \theta_{i+1})), \quad i = 1, ..., n-1,$$

$$p(\theta_n) = \mu C_x(x(\theta_n), \theta_n).$$

Dividing the adjacent first-order conditions renders

$$\frac{p(\theta_{i+1})}{p(\theta_i)} = \frac{C_x(x(\theta_{i+1}), \theta_{i+1}) - C_x(x(\theta_{i+1}), \theta_{i+2})}{C_x(x(\theta_i), \theta_i) - C_x(x(\theta_i), \theta_{i+1})}, \quad i = 1, ..., n-2, \quad (4)$$

$$\frac{p(\theta_n)}{p(\theta_{n-1})} = \frac{C_x(x(\theta_n), \theta_n)}{C_x(x(\theta_{n-1}), \theta_{n-1}) - C_x(x(\theta_{n-1}), \theta_n)}. \quad (5)$$

Intuitively, the equilibrium effort-grade allocations need to be such that the teacher's gains and losses in terms of students' effort changes resulting from a marginal change in the contractual allocations offset each other. Comparing condition (4) with (5), we can notice that the trade-off between gains and losses from a change in the contractual allocations for the two most efficient agents is different from that when changing the allocations for the rest of the agents.

Formally, the right-hand side of (5) can be approximated by

$$\begin{aligned}
&\frac{C_x(x(\theta_n), \theta_n)}{C_x(x(\theta_{n-1}), \theta_{n-1}) - C_x(x(\theta_{n-1}), \theta_n)} = \\
&= \frac{C_x(x(\theta_n), \theta_n)}{C_x(x(\theta_n - \partial\theta), \theta_n - \partial\theta) - C_x(x(\theta_n - \partial\theta), \theta_n)} \approx \\
&\approx \frac{C_x(x(\theta_n), \theta_n)}{-\partial\theta C_{x\theta}(x(\theta_n - \partial\theta), \theta_n)} \approx \\
&\approx \frac{C_x(x(\theta_n), \theta_n)}{-\partial\theta(C_{x\theta}(x(\theta_n), \theta_n) - \partial x C_{x\theta x}(x(\theta_n), \theta_n))}, \quad (6)
\end{aligned}$$

where $\partial x = x(\theta_n) - x(\theta_{n-1})$. The above approximation becomes more accurate with smaller $\partial\theta$.

We observe that the left-hand side of (5), $p(\theta_n)/p(\theta_{n-1})$, tends to 1 with $\partial\theta$ approaching 0, while the right-hand side, in contrast, tends to diverge from 1 when $\partial\theta$ approaches 0 unless the differential in effort level ∂x is negative, but it is precluded by the constraint of the non-decreasing effort x. Therefore, condition (5) cannot hold for a small enough $\partial\theta$ and, in particular, for the continuum of types θ (*i.e.*, at the limit $\partial\theta \to 0$). Hence, the supposition that the solution takes the form of a separating equilibrium, made when constructing implementability constraint (3), is wrong for fine enough distributions of abilities, and in that case there can't be a separating equilibrium among the most efficient agents.

Intuitively, the finding that there is no separating equilibrium among the most efficient agents should not be surprising. Consider the teacher increasing the demanded effort level of the second-most-efficient student from $x(\theta_{n-1})$ to $x'(\theta_{n-1})$ against the corresponding grade increase from $r(\theta_{n-1})$ to $r'(\theta_{n-1}) = 1$, and so that this change is acceptable to the student (otherwise, he would not report truthfully). Comparing the teacher's gains and losses from this change, the loss is the most efficient agent's effort reduction from $x(\theta_n)$ to $x'(\theta_{n-1})$ granting the same reward of 1 [by construction, it must be that $x(\theta_n) > x'(\theta_{n-1})$]. At the same time, the gain accrued to the teacher is not only the increase in the second-most-efficient agent's effort level by $x'(\theta_{n-1}) - x(\theta_{n-1}) > 0$, but actually it is the whole string of follow-up increases in other students' effort levels $x(\theta_i) \rightarrow x'(\theta_i)$, $i = 1, ..., n-2$, made at no extra cost (due to costless rewards) to fill the slack in the incentive-compatibility constraints arisen after the increase in the reward $r(\theta_{n-1})$ to $r'(\theta_{n-1})$. Hence, unless the probability mass of the most efficient type θ_n is big enough (which in the continuous case is possible only for some irregular distribution F for abilities), the teacher's gain from an increase in the second-most-efficient student's effort is larger than the corresponding loss. Eventually, the teacher will increase the expected effort of her students by pooling the most efficient agents, who are set for the highest reward of 1, until their probability mass is big enough to offset the gains and losses described above.

"Bunching at the top" interval

When condition (5) doesn't hold, which occurs, as we showed, for a small enough $\partial\theta$, we proceed by adding up the probability mass of the agents subject to the highest reward of 1 until the updated first-order condition (derived in the same way as before) is fulfilled for some agent type θ_m (let $m > 1$, *i.e.*, there is an interior solution to the problem; the continuous-type case below will be analyzed more generally). The updated condition (5) becomes as follows

$$\frac{P(\theta_m)}{p(\theta_{m-1})} = \frac{C_x(x(\theta_m), \theta_m)}{C_x(x(\theta_{m-1}), \theta_{m-1}) - C_x(x(\theta_{m-1}), \theta_m)} \approx$$
$$\approx \frac{C_x(x(\theta_m), \theta_m)}{-\partial\theta(C_{x\theta}(x(\theta_m), \theta_m) - \partial x C_{x\theta x}(x(\theta_m), \theta_m))},$$

where $P(\theta_m) = \sum_{i=m}^{n} p(\theta_i)$. In case there is no such that θ_m, for which the above condition holds, then the pooling equilibrium interval extends to the whole student type space.

Next, multiplying both sides of the last expression by $\partial\theta$ and taking the limit $\partial\theta \rightarrow 0$ on both sides give the condition for the pooling equi-

librium $[\theta^*, \theta_b]$ among the most efficient agents in the continuous case (and when $\theta^* > \theta_a$):

$$\frac{1 - F(\theta^*)}{f(\theta^*)} = -\frac{C_x(x(\theta^*), \theta^*)}{C_{x\theta}(x(\theta^*), \theta^*)}. \tag{7}$$

Using the assumption that the cost function $C(x, \theta)$ is separable in x and θ with the imposed functional form as in (1), the pooling equilibrium condition (7) becomes free of the effort level term x as below

$$\frac{1 - F(\theta^*)}{f(\theta^*)} = \theta^*. \tag{8}$$

More generally, define $G(\theta) \equiv (1 - F(\theta))/(f(\theta)\theta)$, which is monotonically decreasing in θ due to the monotone hazard rate assumption. If there is a solution to $G(\theta) = 1 : \theta \in [\theta_a, \theta_b]$, then the starting value of the "bunching at the top" interval is $\theta^* = \arg\{G(\theta) = 1\}$, if there is no solution to $G(\theta) = 1 : \theta \in [\theta_a, \theta_b]$, then we set $\theta^* = \theta_a$. More succinctly, the lowest student type subject to the highest reward is determined by $\theta^* = \min\{\theta : G(\theta) \leq 1,\ \theta \in [\theta_a, \theta_b]\}$. All in all, the pooling equilibrium contractual alloaction is $x(\theta) = x^*, r(\theta) = 1$ for every θ in $[\theta^*, \theta_b]$, where the effort level x^* is determined by the remaining dynamics described below.[10]

The remaining dynamics in $[\theta_a, \theta^*)$

After establishing the equilibrium condition for the most efficient students, we look at the remaining interval of agent types $[\theta_a, \theta^*)$ using first-order condition (4), which can expressed as

$$\frac{p(\theta_{i+1})}{p(\theta_i)} = \frac{C_x(x(\theta_{i+1}), \theta_{i+1}) - C_x(x(\theta_{i+1}), \theta_{i+2})}{C_x(x(\theta_i), \theta_i) - C_x(x(\theta_i), \theta_{i+1})} \approx$$
$$\approx \frac{C_{x\theta}(x(\theta_{i+1}), \theta_{i+1})}{C_{x\theta}(x(\theta_i), \theta_i)}.$$

After some further transformations:

$$1 - \frac{p(\theta_i) - p(\theta_{i+1})}{p(\theta_i)} \approx \frac{C_{x\theta}(x(\theta_{i+1}), \theta_{i+1})}{C_{x\theta}(x(\theta_i), \theta_i)},$$

$$C_{x\theta}(x(\theta_i), \theta_i) - C_{x\theta}(x(\theta_i), \theta_i)\frac{p(\theta_i) - p(\theta_{i+1})}{p(\theta_i)} \approx C_{x\theta}(x(\theta_{i+1}), \theta_{i+1}),$$

[10] At this point, we can observe that a pooling equilibrium among the most efficient agents can be obtained with or without the assumption about the log-concavity of the distribution function F, *i.e.*, the monotone hazard rate (see Bagnoli and Bergstrom, 2005), since the interval $[\theta^*, \theta_b]$ cannot be empty because of $G(\theta_b) = 0 < 1$.

and dividing both sides by $\partial\theta$ and rearranging terms give

$$\frac{C_{x\theta}(x(\theta_i+\partial\theta),\theta_i+\partial\theta)-C_{x\theta}(x(\theta_i),\theta_i)}{\partial\theta}\approx$$
$$\approx C_{x\theta}(x(\theta_i),\theta_i)\frac{p(\theta_i+\partial\theta)-p(\theta_i)}{\partial\theta p(\theta_i)}.$$

Taking the limit $\partial\theta\to 0$ on both sides of the last expression renders for any θ (the subscript i, henceforth, is dropped)

$$C_{x\theta x}(x(\theta),\theta)x'(\theta)+C_{x\theta\theta}(x(\theta),\theta)=\frac{f'(\theta)}{f(\theta)}C_{x\theta}(x(\theta),\theta),$$

which is a first-order differential equation for $x(\theta)$. Substituting $g(x)/\theta$ for $C(x,\theta)$ gives

$$g''(x)(-\frac{1}{\theta^2})x'(\theta)+g'(x)\frac{2}{\theta^3}=\frac{f'(\theta)}{f(\theta)}g'(x)(-\frac{1}{\theta^2}), \tag{9}$$

which we can solve for $x(\theta)$:

$$x(\theta)=g^{'-1}(Af(\theta)\theta^2), \tag{10}$$

where A is a constant to be determined. The constraint for the effort function x to be non-decreasing is met, which follows from equations (9), (8), and the monotone hazard rate assumption.[11]

The reward function r is derived from the binding incentive-compatibility constraints ICs, which in the continuous case take the form

$$r'(\theta)=C_x(x(\theta),\theta)x'(\theta),$$

$$r'(\theta)=\frac{g'(x(\theta))}{\theta}x'(\theta),$$

or, using the expression for x as in (10),

$$r'(\theta)=Af(\theta)\theta x'(\theta)=\frac{A^2f(\theta)\theta^2(f'(\theta)\theta+2f(\theta))}{g''(x(\theta))},$$

and, finally, the grade function r can be expressed as

$$r(\theta)=A^2\int_{\theta_a}^{\theta}\frac{f(\tilde{\theta})\tilde{\theta}^2(f'(\tilde{\theta})\tilde{\theta}+2f(\tilde{\theta}))}{g''(x(\tilde{\theta}))}d\tilde{\theta}+B. \tag{11}$$

[11] From condition (8) it follows that $\frac{f(\theta)}{1-F(\theta)}<\frac{1}{\theta}$ for θ in $[\theta_a,\theta^*)$, and from the monotone hazard rate: $\frac{f'(\theta)}{f(\theta)}>-\frac{f(\theta)}{1-F(\theta)}$; the two inequalities together render $\frac{f'(\theta)}{f(\theta)}>-\frac{1}{\theta}$. Given this derived condition and $g'(x)>0$, $g''(x)>0$ (which follow from the assumptions on the cost function C), eq. (9) holds only if $x'(\theta)>0$.

The constant B follows from (11) and the binding IR constraint $r(\theta_a) = g(x(\theta_a))/\theta_a$, which together render

$$B = \frac{g(g^{'-1}(Af(\theta_a)\theta_a))}{\theta_a}. \tag{12}$$

Lastly, the constant A is pinned down through the condition $r(\theta^*) = 1$, which is

$$1 = A^2 \int_{\theta_a}^{\theta^*} \frac{f(\tilde{\theta})\tilde{\theta}^2(f'(\tilde{\theta})\tilde{\theta} + 2f(\tilde{\theta}))}{g''(x(\tilde{\theta}))} d\tilde{\theta} + \frac{g(g^{'-1}(Af(\theta_a)\theta_a))}{\theta_a}. \tag{13}$$

A word of caution should be said here since, using the contract theory terms, we do not mention about a possible shutdown of some inefficient types. Therefore, we should rather replace θ_a with $\theta_* \geq \theta_a$ and provide the condition for the least efficient agent type θ_* considered for non-zero contractual allocations. In the case of a shutdown, the IR_1 constraint would be replaced with that for θ_* resulting in some level shifts for $x(\theta)$ and $r(\theta)$, but the dynamics would remain intact. However, for simplicity we shall ignore this possibility and make only a note that when designing contracts with non-pecuniary rewards the incidence of a shutdown is less likely than when designing contracts with pecuniary rewards. For instance, with the cost function $C(x, \theta) = x^2/(2\theta)$, the teacher will never exclude any of student types irrespective of her beliefs about the distribution of student abilities, which is, generally, not true if we have a standard principal-agent model with pecuniary rewards and the above cost function C.

Proposition 1 below summarizes the solution to Program 1.

Proposition 1 *Given the assumptions of the model, the solution to Program 1—the reward function r and the effort function x—is characterized by:*

- *for every agent type θ in $[\theta^*, \theta_b]$, where $\theta^* = \min\{\theta : \frac{1-F(\theta)}{f(\theta)\theta} \leq 1, \theta \in [\theta_a, \theta_b]\}$, the uniform contract applies: $r(\theta) = 1$, $x(\theta) = x(\theta^*)$, where $x(\theta^*)$ is defined by* (10)*;*
- *for every θ in $[\theta_a, \theta^*)$ the optimal contract $\{x(\theta), r(\theta)\}$ is defined by (10), (11), (12)and (13).*

4 Main Findings and Discussion

Here, we provide an intuitive account of the obtained solution to the agency problem with costless non-pecuniary rewards. Simultaneously, in this and subsequent sections, we shall try to relate the predictions of the model to the empirical evidence relevant to the examined framework.

4.1 Compression of Ratings

One of the main theoretical results of the model is that in a principal-agent model with costless non-pecuniary rewards we inevitably obtain a pooling equilibrium for at least some of the most efficient agents, which is in stark contrast to the standard model with pecuniary rewards, where the "no distortion at the top" property typically holds.[12]

Proposition 2 *For an agency problem with costless rewards, the "no distortion at the top" property ceases to hold.*

The proof of this result has already been provided in the previous section, when we showed that a uniform contract applies to all agent types from the non-empty interval $[\theta^*, \theta_b]$. Moreover, the above proposition is silent about the reward function r to be bounded, for it has no impact on the result. As we have already argued in footnote 5, without an upper bound on the reward function the uniform contract would apply to all agent types.[13]

Nor is the result in Proposition 2 influenced by the specific functional form of the convex cost function C that is assumed in (1), for the result primarily hinges on the universally present term $(1 - F(\theta))/f(\theta)$ in (7). Should we solve Program 1 for any teacher's utility function V, we would get a condition equivalent to (5) as below

$$\frac{p(\theta_n)}{p(\theta_{n-1})} = \frac{V_x(x(\theta_{n-1}))}{V_x(x(\theta_n))} \frac{C_x(x(\theta_n), \theta_n)}{C_x(x(\theta_{n-1}), \theta_{n-1}) - C_x(x(\theta_{n-1}), \theta_n)}. \qquad (14)$$

The above expression is identical to (5) except for the term involving the derivative of V in x, which compares the marginal utility from the efforts $x(\theta_{n-1})$ and $x(\theta_n)$ accruing to the principal. If the function V is concave in x, then the pooling equilibrium interval extends compared to that in the linear case. The teacher attempts to boost the incentives of low performers even further despite that it comes at the expense of lower effort levels demanded from high-ability students. In particular, since the expression $V_x(x(\theta_{n-1}))/V_x(x(\theta_n))$ is greater than 1, the argument following the derived expression (6) is reinforced. However, with

[12] Though, recently, there have also been papers arguing the opposite. For example, Levin (2003) presents a multiperiod adverse selection framework with observable but not verifiable effort levels, at which the "no distortion at the top" property may not hold.

[13] In the principal-agent model with costly rewards, imposing an upper bound on the reward function does not lead to a pooling equilibrium among the most efficient agents as long as this constraint is not binding, *i.e.*, the upper bound is big enough. However, in our model, irrespective of the size of the upper bound, we do obtain a pooling equilibrium among the most efficient agents.

a convex utility function V the pooling equilibrium interval shrinks, because now the teacher cares more about high performers than about low performers, but the interval still remains non-singleton unless the function V is too steep for some feasible effort level x. However, the latter possibility is excluded by requiring the teacher's utility function V being "less convex" than the student's cost function C.[14]

As has already been suggested, the pooling-equilibrium interval may stretch out to comprise the whole agent-type space.[15] Having this result in mind, we can address the phenomenon of the compression of ratings, mentioned in the introduction. On the above grounds, it may not be optimal for a supervisor to differentiate much among her employees when evaluating their performance. The attempts to distinguish the most productive agents are achievable only by suppressing the motivation of less productive agents, which may have a larger adverse effect on aggregate output than the achieved higher contribution from the most efficient agents. Hence, the "leniency bias" rating scheme could be, in fact, optimal, which is to give the same appraisal to every agent, provided some minimum level of effort is exerted.[16]

Specifically, thinking of a job performance appraisal by ratings and trying to link it to the model developed above, it remains true that a manager wants her employees to put more effort and is eager to motivate them in different ways. However, differently from a teacher-student relationship, the evaluation of job performance in terms of only grades or ratings not always provides an adequate motivation for an employee. Typically, a rating scheme is linked to the employees' pay scheme, with a higher rating implying a higher pay, which, of course, does not fit the definition of "costless non-pecuniary rewards". However, having inquired into the inner workings of rating-pay schemes[17], in many instances we can still think of job performance appraisals in terms of a principal-agent model with costless rewards. Commonly, employees' performance

[14]The above discussion shows that changing teachers' incentives can have direct implications on their student grading schemes. Hence, through pay-for-student-performance incentive programs it could be possible to align teachers' incentives with those of the social planner.

[15]For this to occur in our model,it is enough, for instance, to have a quadratic cost function $C = x^2/(2\theta)$, the uniform distribution for the agent type θ, and that $\theta_a \geq \theta_b/2$.

[16]The above explanation is in stark contrast to explanations from psychological literature on subjective evaluation, a common example of which is given in Prendergast (1999), footnote 34: "An obvious reason for this [leniency bias] is that it is simply unpleasant for supervisors to offer poor ratings to workers, so they avoid this pain."

[17]I largely owe the following discussion to Ailko van der Veen, who shared with me his experience from working at the banking industry after the presentation of the early version of this paper at the ENABLE workshop in Amsterdam, 2006.

is evaluated by their line managers, *i.e.*, by low rank managers, whose own incentive scheme does not always internalize fully the payroll cost resulted from their evaluations of employees' performance, but it does depend on the aggregate performance of the employees in charge (say, through bonuses at the end of the year).[18] Therefore, we may obtain a situation when a line manager would face a soft-budget constraint when rewarding her employees with ratings (costless to her, and on a finite scale), but would also try to motivate the highest expected effort from them. The possibility described, then, falls within our model, but, of course, for its specific nature it needs to be studied separately, and here it serves only as a potential alternative application of the model.

4.2 Mismatch between Grades and Abilities

Consider two classes of students with a general belief that students from the first class are more able than those from the second class. In mathematical terms, let student types from the first and second classes be random variables Θ_1 and Θ_2 with twice differentiable distribution functions F_1 and F_2 defined on the same support $[\theta_a, \theta_b]$, respectively. We understand by the general belief that the first class is more able than the second class in terms of the hazard-rate dominance:

$$h_1(\theta) \leq h_2(\theta) \text{ for every } \theta \text{ in } [\theta_a, \theta_b],$$

where $h_i(\theta) = f_i(\theta)/(1 - F_i(\theta))$ is the hazard rate of a random student type Θ_i, $i = 1, 2$. In other words, Θ_2 is smaller than Θ_1 in the hazard rate order, which can also be expressed as $E_1(\theta \mid \theta > \bar{\theta}) \geq E_2(\theta \mid \theta > \bar{\theta})$ offering a more intuitive account of the suggested dominance condition. Given the same signal that a student's ability is at least $\bar{\theta}$, the teacher of the first class will hold higher expectations of the actual student ability than will the other teacher.

Proposition 3 *The lower expectations the teacher holds about her students' abilities, the more lenient she should be in grading them.*

More formally, let $\{x_1(\theta), r_1(\theta)\}$ and $\{x_2(\theta), r_2(\theta)\}$ be the solutions to Program 1 with the distributions F_1 and F_2 for two classes of students, respectively, such that the hazard rates satisfy $h_1(\theta) \leq h_2(\theta)$ for

[18] This situation is more likely to happen when the output of a particular production unit cannot be immediately assessed if it only contributes to the final product, therefore, it is complicated to design the incentive schemes for managers that would link labor costs to the produced output (to alleviate which, there is a new practice of establishing separate profit centers within large companies; see Frey and Osterloh, 2002).

every θ in $[\theta_a, \theta_b]$ (with some additional stochastic-dominance qualification specified in the proof below). Then, the relationship between the resulting optimal grade allocations is

$$r_1(\theta) \leq r_2(\theta) \quad \text{for every student type } \theta,$$

which is in words—the more pessimistic the principal is about her agents, the more generous she should be in motivating them.

Proof. With reference to Proposition 1, define $G_i(\theta) = (1-F_i(\theta))/\theta f_i(\theta)$, and let $\theta_i^* = \min\{\theta : G_i(\theta) \leq 1, \theta \in [\theta_a, \theta_b]\}$, $i = 1, 2$. Evidently, if $f_1(\theta)/(1 - F_1(\theta)) \leq f_2(\theta)(1 - F_2(\theta))$ for every θ, then $G_1(\theta) \geq G_2(\theta)$, leading to $\theta_1^* \geq \theta_2^*$. In words, the pooling-equilibrium interval $[\theta^*, \theta_b]$, for which the highest reward of 1 is granted, is larger for the class of less able students. In particular, we have that $r_2(\theta) = 1 \geq r_1(\theta)$ for $\theta \in [\theta_2^*, \theta_b]$. To prove that $r_1(\theta) \leq r_2(\theta)$ for the student types θ in $[\theta_a, \theta_2^*)$, we need to strengthen the stochastic dominance assumption for the support interval in question. In particular, we need the dominance in the likelihood-ratio order, *i.e.*, $f_1(\theta)/f_2(\theta)$ is increasing in $\theta \in [\theta_a, \theta_2^*]$.[19] Since the effort level $x_2(\theta_2^*)$ must be at least as large as $x_1(\theta_2^*)$, which stems from the second teacher's incentive to expand the pooling equilibrium even further than the first teacher is inclined to do, then from eq. (10) it follows that $A_2 f_2(\theta_2^*) \geq A_1 f_1(\theta_2^*)$. Given the increasing likelihood ratio $f_1(\theta)/f_2(\theta)$, we obtain that the inequality $A_2 f_2(\theta) \geq A_1 f_1(\theta)$ holds for every θ to the left from θ_2^*, which from eq. (10) renders $x_2(\theta) \geq x_1(\theta)$ for every θ in $[\theta_a, \theta_2^*)$, which subsequently leads to $r_2(\theta) \geq r_1(\theta)$. ■

Figure 1 below illustrates Proposition 3. The two diagrams on the left show the distributions for student types that the teachers of the first class (top-left) and second class (bottom-left) hold. Respectively, the diagrams on the right depict the optimal effort-grade allocations for the two classes. As discussed above, the teacher of the first class offers the highest grade of 1 to fewer students but against a higher cost (effort) than the teacher of the second class does. In general, comparing the grade-effort allocations in both classes, we are to observe that the teacher of the first class with relatively more able students offers a steeper grade-effort schedule. The reason for this is that the first teacher in her attempts to extract more effort from good students makes sure that they are not lured to take less effort, for it will result in a substantially smaller grade. Differently, the teacher of the second class focuses on less able but more numerous students and attempts to extract more effort from them leaving more able students with high information rents.

[19] It hasn't been done before (just a qualification was made) for the reason that the hazard-rate order dominance is sufficient to obtain the difference in the pooling-equilibrium intervals, which is the driving result of the proposition.

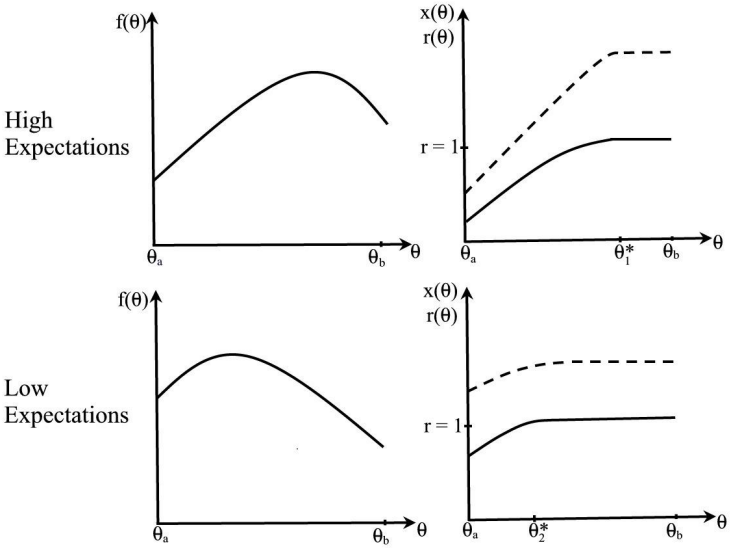

Figure 1. Teacher's belief f, and optimal grading scheme: grade r (solid line) and effort x (dashed line).

In this light, it should not be surprising from our analysis perspective that, given differences in the classes of students, some teachers turn out to be more generous than others, leading to the phenomenon of the mismatch between grades and abilities with direct implications to grade inflation (for an extensive study into this problem, see Johnson, 2003). Arguably, if students taking, say, mathematics classes are more capable than those students opting for less involved classes, then, as the model predicts, we should observe fewer highest grades among the mathematics students than among those taking an easier class. Hence, within our developed framework we can obtain a low correlation between grades and abilities (as there is ample empirical evidence on that, which is explored in the following section), but that could be the outcome of the optimal incentive scheme design, and not necessarily the outcome of some teachers' rent-seeking behavior, as sometimes is suggested (*e.g.*, Johnson, 1997).

Of course, the normative side of the issue that differences in grading standards may create perverse incentives to some students has to be separately examined, however, it needs to be stressed that the cause of those differences could be that teachers rationally take into account variation in student abilities when designing grading rules.

5 Empirical Evidence

The theoretical predictions of the model, in particular that in Proposition 3, seem to be empirically testable, for the necessary data needed for this purpose such as student grades and their ability proxy (like their performance on university entry exams or Scholastic Aptitude Test [SAT] scores) should be available at any university. Then, we would need, roughly speaking, to compare grading patterns at classes with different student ability distributions. However, and not surprisingly, there have been a number of empirical studies of the kind in the special literature of educational measurement (*e.g.*, in academic periodicals such as *Journal of Educational Measurement* or *Educational and Psychological Measurement*). Most importantly, those studies without an exception do report the results that are fully in line with the model's predictions: studied fields with lower ability students as compared with those of higher ability students employ less stringent grading criteria. Despite that many of those studies are fairly comprehensive in empirical matters, they lack any rigorous theoretical explanation for this phenomenon, hinging mainly on the level of intuition or reference to similar phenomena from the adaptation-level theory in psychological literature. In what follows, we attempt to review closely some of the empirical studies comparing grading standards throughout time and among different fields, and to show that the model developed here proves helpful in explaining the observed empirical evidence.

Aiken (1963) is one of the first empirical studies that suggest that the grading behavior is dictated by the quality of students in the current class and not by some absolute invariant standards. Aiken (1963) presents time-series evidence from the Woman's College of the University of North Carolina that could imply that with more able students in a class (as measured by their SAT scores and high-school rankings) teachers tend to apply more stringent grading standards. As for the theoretical explanation of this finding, the study just briefly mentions that it conforms with the adaptation-level theory or central tendency phenomenon, which is basically about the tendency of supervisors to evaluate the performance of the supervised in relative terms rather than in absolute ones.

A much more comprehensive study Goldman and Widawski (1976) firstly point out the weaknesses of previous studies on grading patterns for their using the total grade point average (GPA) as the criterion of grading standards. As they justly argue, GPAs are neither perfectly comparable throughout time nor among individual students because of a possibly different composition of courses included to compute grade averages. To remedy that, Goldman and Widawski (1976) employ a

between-subjects design aimed at making grades' comparisons more effective. They compute an index of grading standards using pairwise comparisons of grades in 17 major fields at the University of California, Riverside, from a random sample of 475 students. In particular, they perform the comparison of grading standards in one class (say, psychology) against those in another class (say, biology) by computing the difference in average grades of only those students who took both classes. After obtaining differentials in grading standards between any two classes (from 17 classes available in their study), they construct an index of grading standards for each class, which is an average of all the differentials between that particular class and the rest of the classes. Finally, they correlate the computed indices of grading standards with the average scores on the verbal and mathematical portions of the SAT test and high-school GPAs (*i.e.*, student ability proxies) of all the students majoring in those 17 classes. The main empirical finding in Goldman and Widawski (1976) is that the constructed index of grading standards correlates highly in a negative direction with student ability proxies. In other words, they conclude that professors in a field containing more able students tend to grade more stringently than do professors in fields with lower ability students. As a result of that, they find that the past performance and abilities of students account for only slightly more than 50 percent of the variance in grades, and suggest introducing some grade adjustment mechanism to make grades more informative of students' true abilities. Again, as for giving an explanation of the obtained empirical results, they restrict the argument just to making a reference to the adaptation-level theory that people are judged in comparison to their peers.

In a similar study Goldman and Hewitt (1975), besides presenting the empirical results (which draw the same conclusions about grading behavior as in the studies mentioned above), there is also a more elaborate theoretical explanation for the obtained results. The authors believe that the antecedents (*e.g.*, student ability levels, work habits, *etc.*) and consequences (grading standards) of college grading are inextricably tied together by a personal characteristic of college instructors. This characteristic is the phenomenon of adaptation level, and it is so pervasive among college instructors and perhaps people in general, Goldman and Hewitt (1975) continue, as to be considered an almost inevitable factor in college grading process. Consequently, through that personal characteristic link grading standards would be partly determined by the ability level of the student population. However, along the lines of our model, developed above in the text, this personal characteristic, as envisaged by Goldman and Hewitt (1975), is not some intrinsic feature of human

behavior but rather the outcome of optimal behavior.

A decade later, Strenta and Elliott (1987) replicate the study of Goldman and Widawski (1976) using data from a different institution, Dartmouth College, just to find that the differential grading standards exist in the same magnitude and in roughly the same order. Hence, Strenta and Elliott (1987) argue that it remains the case that students with higher SAT scores tend to major in fields with more rigorous grading standards, and that factors attracting more talented students result in their being graded harder. (Though, we would say that because more talented students are attracted to some particular classes, the professors of those classes tend to grade them more stringently, which is to say, optimally.) As before, Strenta and Elliott (1987) continue arguing that these differential grading standards serve to attenuate the correlation between the GPAs and SAT scores of the students. However, they also show that the correlation sizably increases if GPAs are adjusted by accounting for differences in departmental grading standards. Finally, a similar study conducted in Duke University (Johnson, 2003) confirmed the conclusions about systematic differences in grading standards of the previous studies.

Concerning the normative side of the discussed differential grading standards, there have been a number of papers proposing grade adjustment mechanisms (see, *e.g.*, Johnson, 1997) in order to make grades more informative of students' actual abilities. Without going into the details of this literature, it is worth noticing that, typically, those papers tend to assume that the true reason for differential grading standards lies with some personal features of the instructor (*e.g.*, the adaptation level, unwillingness to spend office hours on dealing with students' complaints about low grades, *etc.*). Therefore, the proposed grade adjustment mechanisms would attempt to correct for presumed instructor-specific factors failing to recognize the possible endogeneity of those factors, which could lead the mechanism astray from the projected goals.

6 Conclusion

In this paper, we solve for the optimal contract in an agency problem featuring costless non-pecuniary rewards, and apply the obtained results to provide alternative explanations for the compression of ratings and mismatch between students' abilities and grades. We argue that in equilibrium the variation in assigned rewards can be coarser than the underlying distribution for abilities. In particular, unless the principal is very optimistic about the overall distribution of agent abilities, to set uniform incentives for all agents conditional on their achieving some prespecified minimum standard can constitute an optimal contract. Specifically for

student grading schemes, if the teacher's goal is to induce her students to study as hard as possible, we should observe higher grades in classes with fewer able students. Importantly, the existing empirical evidence strongly supports the predictions of the model presented in this paper lending validity to the chosen modeling technique.

Therefore, the proposed framework could be potentially used as the "microfoundations" of student grading or job performance appraisal to analyze other related problems. For instance, if the model were made dynamic, then we could potentially look into the phenomenon of grade inflation over time or how to design the evaluation process to reduce the observed inefficiencies (*i.e.*, coarse grading or rating schemes) of a static relationship. Further research could be done on studying the implications on student effort-grade allocations after the introduction of incentives for teachers or on developing grade-adjustment mechanisms to make the intercomparison of grades between various classes, departments or schools possible.

References

[1] Aiken, L. (1963). "The grading behavior of a college faculty." *Educational and Psychological Measurement*, 23, 319–322.

[2] Akerlof, G. (1982). "Labor Contracts as Partial Gift Exchange." *Quarterly Journal of Economics*, 97, 543–569.

[3] Arcidiacono, P. (2004). "Ability sorting and the returns to college major." *Journal of Econometrics*, 121, 343–375.

[4] Atkinson, A., Burgess, S., Croxson, B., Gregg, P., Propper, C., Slater, H., and Wilson, D. (2004). "Evaluating the Impact of Performance-related Pay for Teachers in England." *CMPO working paper* No. 04/113.

[5] Bagnoli, M. and Bergstrom, T. (2005). " Log-concave probability and its applications." *Economic Theory*, 26, 445–469.

[6] Benabou, R. and Tirole, J. (2003). "Intrinsic and Extrinsic Motivation." *Review of Economic Studies*, 70, 489–520.

[7] Berg, J., Dickhaut J. and McCabe, K. (1995). "Trust, Reciprocity, and Social History." *Games and Economic Behavior*, 10, 122–142.

[8] Bolton, P. and Dewatripont, M. (2004). *Contract Theory.* MIT Press.

[9] Brennan, G. and Pettit, P. (2004). *The Economy of Esteem: An Essay on Civil and Political Society.* Oxford University Press.

[10] Dubey, P. and Geanakoplos, J. (2005). "Grading in Games of Status: Marking Exams and Setting Wages." *Cowles Foundation Discussion Paper* #1544.

[11] Falk, A. and Kosfeld, M. (2006). "Distrust – The Hidden Cost of

Control." *American Economic Review*, 96, 1611–1630.
[12] Frey, B. S., and Osterloh, M. (2002). *Successful management by motivation: Balancing intrinsic and extrinsic incentives*. Berlin, Heidelberg, New York: Springer.
[13] Fudenberg, D. and Tirole, J. (1991). *Game Theory.* MIT Press.
[14] Goldman, R. and Hewitt, B. (1975). "Adaptation-level as an explanation for differential standards in college grading." *Journal of Educational Measurement*, 12, 149–161.
[15] Goldman, R. and Widawski, M. (1976). "A within-subjects technique for comparing college grading standards: implications in the validity of the evaluation of college achievement." *Educational and Psychological Measurement*, 36, 381–390.
[16] Guesnerie, R. and Laffont, J.J. (1984). "A Complete Solution to a Class of Principal-Agent Problems with an Application to the Control of a Self-Managed Firm." *Journal of Public Economics*, 25, 329–69.
[17] Johnson, V.E. (1997)."An Alternative to Traditional GPA for Evaluating Student Performance." *Statistical Science*, 12, 251–278.
[18] Johnson, V.E. (2003). *Grade Inflation: A Crisis in College Education*. New-York: Springer.
[19] Lavy, V. (2002). "Evaluating the Effect of Teachers' Group Performance Incentives on Pupil Achievement." *Journal of Political Economy*, 110, 1286–1317.
[20] Lazear, E. (2003). "Teacher Incentives." *Swedish Economic Policy Review*, 10, 179–214.
[21] Levin, J. (2003). "Relational Incentive Contracts." *American Economic Review*, 93, 835–857.
[22] Mirrlees, J.A. (1971). "An Exploration in the Theory of Optimum Income Taxation." *Review of Economic Studies*, 38, 175–208.
[23] Murphy, K. and Cleveland, J. (1995). *Understanding performance appraisal: Social, organizational and goal-oriented perspectives*. Newbury Park, CA: Sage.
[24] Ostrovsky, M. and Schwarz, M. (2005). "Equilibrium Information Disclosure: Grade Inflation and Unraveling." *Working Paper, Harvard Institute of Economic Research.*
[25] Prendergast, C. (1999)."The Provision of Incentives in Firms." *Journal of Economic Literature*, 37, 7–63.
[26] Strenta, A. and Elliott, R. (1987). "Differential grading standards revisited." *Journal of Educational Measurement*, 24, 281–291.
[27] Sliwka, D. (2007). " Trust as a Signal of a Social Norm and the Hidden Costs of Incentive Schemes." *American Economic Review*, 97, 999–1012.