



Institute for Empirical Research in Economics  
University of Zurich

Working Paper Series  
ISSN 1424-0459

---

Working Paper No. 487

**Optimal Grading**

Robertas Zubrickas

May 2010

---

# Optimal Grading\*

Robertas Zubrickas<sup>†</sup>

April 6, 2010

## Abstract

Assuming that teachers are concerned with human capital formation and students — with ability signaling, in this paper we model a teacher-student relationship as an agency problem with conflicting interests. In our model, the teacher elicits effort from the student rewarding for it with a grade, the utility of which to the student is an ability signal inferred by the job market. In the event that the job market does not observe individual teachers' grading practice, teachers find grades as costless rewards and optimally choose to be lenient in grading. As a result, “the problem of the commons” of good grades emerges leading to the depreciation of grading standards and grade inflation. The prediction of the model that the lower the expectations the teacher holds about her students' abilities, the flatter the grading rules she sets up is empirically supported.

*Keywords:* Principal-agent model, teacher-student relationship, costless rewards, grading rules, mismatch of abilities and grades, grade inflation, teacher incentives.

*JEL codes:* C70, D82, D86, I20.

## 1 Introduction

In this paper, we approach a teacher-student relationship as a principal-agent model with conflicting interests. Arguably, a teacher's goal is to pass on knowledge to students—the more, the better—while the students care for the teacher's assessment of their performance—grades. A justification of this conflict of interests could be the dichotomy of the role of education: human capital formation versus job market signaling.<sup>1</sup> Therefore, an interpretation of our modeling framework to be presented is that the teacher is more concerned with the human capital formation side, while the students—with the “ability signal” their

---

\*I would like to thank Tore Ellingsen, Christian Ewerhart, Drew Fudenberg, Magnus Johannesson, Karl Wärneryd, and seminar participants at the University of California, Santa Barbara.

<sup>†</sup>University of Zurich, IEW, Winterthurerstrasse 30, CH-8006 Zurich, Switzerland. E-mail: robertas.zubrickas@uzh.ch.

<sup>1</sup>See, e.g., Bedard (2001).

accomplished education carries along, treating knowledge obtained as abstract and useless outside academia. We are going to show that the grading patterns widely observed in practice—a mismatch or low correlation between students’ grades and their abilities and coarse grading with implications to grade inflation—can be the outcomes of our agency problem, as studied from an individual teacher’s perspective.

We consider a grading rule—assigning grades to exam scores—as “a contract” that the teacher offers her student. With the help of grades, the teacher aims to elicit costly learning effort from—equivalently, to pass on knowledge to—her student, who comes from a population of students with disparate abilities for the subject taught. The teacher has a technology—the exam test—that allows her to assess the student’s knowledge level attained, on which she conditions grades to be rewarded. We assume that a knowledge level attained is in a direct and deterministic relationship with the student’s learning effort elicited (conditional on his ability type). The main constraint that we impose on the teacher is that her grading rule needs to be incentive compatible, i.e., grades cannot be conditioned on the student’s ability level, which is his private information.<sup>2</sup>

The key feature of this model is that grades are costless rewards for the teacher to give but of a value to the student. We separately consider two sources of value of grades to the student: 1) nominal—a grade is of a value on its own—and 2) relative—the value of a grade comes in the form of an ability signal inferred by the job market. The latter interpretation of the value of grades is in the tradition of treating education as job market signaling (Spence (1973)), but the former is not a big departure from this tradition either. A student may care for his grade at every class he takes not just because of its immediate job-market-signaling value but for its contribution toward his final grade average at the end of his study program, which later on will serve as an ability signal. Besides, external criteria of academic performance—used, e.g., for scholarship application purposes—are usually set in nominal grades.

The main distinction between the two approaches to modeling the utility of grades can be expressed in terms of the scarcity of grades available to the teacher. In the first case, the (*ex ante*) utility of a grade is independent of the frequency of the grade given in the class—i.e., there is no scarcity of grades—while in the second case, where the utility of a grade is equal to the expected ability inferred by the job market, grades are scarce. However, the two approaches turn similar if we assume that the job market cannot identify the teacher’s grading rule applied to grade the student and infers the student’s ability from his grade based on its own perception of grading standards. Then,

---

<sup>2</sup>If a hidden-information framework seems restrictive—arguably, teachers have access to students’ previous records and can learn about their abilities—then, alternatively, we could require that the teacher cannot discriminate among her students by applying ability-specific grading rules (i.e., the same exam score needs to result in the same grade irrespective of the student’s ability level). With this alternative formulation, the optimization problem would remain intact as in the case with the hidden-information framework adopted, and, therefore, the latter is retained for its link with the existing literature.

the teacher essentially finds grades as costless and abundant rewards to give (but with implications to the deterioration of grading standards as later discussed).<sup>3</sup>

All in all, we study a teacher-student relationship in the framework of a single-agent model with hidden information. With the assumption that the teacher’s objective is to maximize the student’s expected knowledge under the incentive compatibility constraint, the goal of this paper is to analyze the properties of the grading rule “optimally” designed by the teacher.

We obtain the following predictions about optimal grading patterns. In the event when students value grades nominally, we observe the teacher pooling student ability types for the highest grade. In other words and more generally, “the no distortion at the top” property does not hold when the teacher (principal) can costlessly reward the student (agent). This result arises because the gain from distorting incentives for students of the highest ability outweighs the corresponding loss (unlike in the standard model with costly transfers). For the case with the relative value of grades, if the student’s learning costs are not too high (i.e., the effort cost function is not too convex), then the teacher designs a grading rule that perfectly screens student ability types.<sup>4</sup> However, if the teacher’s grading rule cannot be distinguished by the job market, we again obtain pooling student types for the highest grade (for the same reasons as in the model with nominal grades). In addition, our comparative statics analysis shows that if the teacher holds low expectations about her students’ abilities, then she should apply more lenient grading standards (in order to elicit on average higher effort levels), and *vice versa*. As a result, this can lead to heterogeneous distributions of grades among classes different in student abilities—in particular, to a mismatch and low correlation between students’ grades and their abilities.

Significantly, the existing empirical evidence strongly supports the findings of the model, lending credibility to our chosen modeling strategy of a teacher-student relationship. With regard to our comparative statics result, Goldman & Widawski (1976) report a negative correlation between students’ Scholastic Aptitude Test scores (which could be seen as a proxy measure of students’ abilities) and the grading standards in the classes the students were majoring in. According to this study (conducted at University of California, Riverside), the negative correlation observed is due to the fact that professors in a field consisting of 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 & Elliott (1987)) and at Duke University (Johnson (2003)). Later in the text,

---

<sup>3</sup>In our model, the student is indifferent about the resultant distribution of grades in the class and, accordingly, his class ranking as long as it does not affect the ability-signaling value of his grade. We do not model that the (*ex ante*) utility of a grade may also come in the form of status or class ranking, unlike in Dubey & Geanakoplos (2009), a related paper discussed later, Moldovanu *et al.* (2007), or Besley & Ghatak (2008).

<sup>4</sup>This result is in line with the results from related models modeling non-pecuniary rewards such as status incentives; see Moldovanu *et al.* (2007); Besley & Ghatak (2008).

we discuss this empirical evidence more thoroughly.

Our model can also offer an insight into the grade inflation phenomenon—“an ongoing rise in grade point averages without an accompanying rise in student ability or effort.”<sup>5</sup> If we take a dynamic perspective that over time, due to an increasing number of university openings available (McKenzie & Tullock (1981)), the distribution for abilities of students enrolled become gradually skewed to the end of lower-ability types<sup>6</sup>, then, according to our model with nominal grades, the teacher should become more lenient with grading. The reason for it is to extract more effort from increasingly numerous lower-ability types (even at the expense of distorting incentives for high-ability students even further). Along the same lines, with the number of educational institutions increasing, an individual teacher’s grading practice bears an increasingly smaller weight on the job market’s perception of grades as ability signals. Then, similarly to the case with nominal grades, in the model with relative grades teachers may find grades to be costless rewards to give and tend to exploit good grades to their benefit. As a result, “the problem of the commons” of good grades arises, which leads to the deterioration of grading standards and grade inflation in the end. Hence, we argue that grade inflation can arise from teachers’ optimal response to changes in the environment they work at.<sup>7</sup>

The key driver of our results obtained is the costlessness of rewards. The proposed refinement that, unlike the agent, the principal is indifferent to a transfer between them is by no means new in the contract theory literature. It was formally studied, for example, in Guesnerie & 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, where with the former preferences the principal’s utility does not depend on a transfer, while with the latter (conventional) preferences it does. In their study, however, the “type A” preferences are primarily used to analyze a social planner’s problem of social welfare maximization. There, a transfer between the social planner (principal) and the agent is equivalent, figuratively speaking, to distributing money between two pockets of the same jacket, leaving the social welfare intact. Therefore, the framework of Guesnerie & Laffont (1984) does not apply to the problem studied here. In our model, the principal is, in fact, more of “type B”, i.e., she cares only about her own utility but does not pay for

---

<sup>5</sup>Dickson (1984). For more on grade inflation, see, e.g., Sabot & Wakeman-Linn (1991) or Kuh & Hu (1999) and other references cited therein.

<sup>6</sup>... which is consistent with the observation of declining college entrance test scores (Wilson (1999)).

<sup>7</sup>The recent papers by Chan *et al.* (2007) and Ostrovsky & Schwarz (Forthcoming) also show how grade inflation can arise in equilibrium. These papers model student grading as a signaling (cheap-talk) game between schools and the job market, and they show that grade compression (and, accordingly, grade inflation) can be an equilibrium disclosure of information by schools about their students’ abilities that maximizes their expected wages (unlike expected learning effort as in our model). However, these papers do not attempt to explain the other empirical observation about grading patterns discussed, namely, the low correlation between students’ grades and their abilities. Later in the paper, we use their argument when discussing policy applications of restricting grading rules applied by teachers.

motivating the agent.

Nor does this paper stand alone in designing optimal grading rules from the perspective of a principal-agent model.<sup>8</sup> Dubey & Geanakoplos (2009) target the question what grading scale (finer or coarser) a teacher should use in order to induce a higher effort from her students. They approach the problem of optimal grading schemes from a perspective different from ours: they model a teacher-student relationship as a game of status with stochastic output similar to a tournament. In their multiple-agent model, a student’s utility of a grade depends on his or her class ranking, i.e., status, resulting from the grade rewarded (but not on a grade *per se* even if a grade carries the same ability signal irrespective of distribution for grades in the class). In addition, in their model the teacher aims to incentivize all her students to put in maximal effort rather than to obtain the highest expected effort. Given these modeling differences, we draw different conclusions about optimal grading schemes. Dubey & Geanakoplos (2009) 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.<sup>9</sup> Our model predicts that teachers should apply coarse grading schemes, but only when they can “costlessly” reward the student (e.g., when an individual teacher’s grading practice cannot affect the perception of the job market about the value of grades), otherwise they should apply a fine grading rule. In addition, in the case with costless grades, we do not find “pyramiding” to be an optimal grading rule, especially, when there is a large mass of less able students in the class.

The remainder of the paper is organized as follows. In Section 2, we introduce a modeling framework. In Section 3, we solve the model for the case when students value grades nominally, and in Section 4—for the case with the relative value of grades. In Section 5, we discuss the main findings of the model(s) presented and relate them to the grading patterns observed in practice; there, we also discuss policy applications of the model(s). In Section 6, we review existing empirical evidence on mismatch between students’ grades and their abilities. The last section concludes the study.

## 2 Framework

There is a teacher teaching her class, made of one student, a particular subject. The teacher’s goal is to pass on knowledge in her subject to the student. She has a technology—the exam test—that allows her to assess the student’s knowledge level attained from his

---

<sup>8</sup>However, 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 part of a more global game involving potential employers or university administration (see Ostrovsky & Schwarz (Forthcoming) or Chan *et al.* (2007)). At the same time, more research has been done on the empirical side of the problem (see Johnson (2003)).

<sup>9</sup>Moldovanu *et al.* (2007) makes a similar prediction as well.

test score  $x \in [0, \bar{x}]$ . (The upper bound on test scores,  $\bar{x}$ , is large enough to allow for an interior solution.) We assume that the teacher’s technology is perfect in the sense that there is a deterministic and direct relationship between test scores and knowledge levels, both of which, therefore, are used synonymously throughout. Achieving a test score  $x$  comes to the student at the effort cost of  $C(x, \theta)$ , where parameter  $\theta$  is the student’s privately known ability level for the subject studied, distributed in the population according to a common prior distribution  $F$  over the ability space  $\Theta = [\underline{\theta}, \bar{\theta}]$ ,  $\bar{\theta} > \underline{\theta} > 0$ . The properties of the effort cost function  $C$  are  $C_x > 0$ ,  $C_{xx} > 0$ ,  $C_\theta < 0$ , and  $C_{x\theta} < 0$ .

The student selects the class for exogenous reasons (e.g., it is compulsory in his curriculum). The reward pursued by the student from participating in the class is the teacher’s assessment, grade  $r \in [0, \bar{r}]$ , of his class performance—knowledge attained. The upper bound on grades,  $\bar{r}$ , is assumed to be institutionally preset, as is the very grading framework, i.e., the assessment of student performance needs to be done in the form of grades only. As already said in the introduction, we separately consider two sources of value of grades to the student: 1) a grade is of a value on its own (Section 3. Nominal Value of Grades) and 2) the value of a grade comes in the form of an ability signal as inferred by the job market (Section 4. Relative Value of Grades). The exact forms of the student’s utility function of a grade are given in the corresponding sections.

The teacher and student’s relationship develops as follows. First, the teacher sets up a grading rule that assigns grades  $r$  to test scores  $x$ . Then, after observing the grading rule, the student decides on a learning effort  $C(x, \theta)$  to achieve the test score  $x$  rewarding the grade  $r$ . The teacher’s objective is to maximize her expected utility, which increases in the student’s knowledge—equivalently, in his test score—and the student’s objective is to maximize his utility of a grade less the effort cost spent to obtain it.

Here, we impose some further structure on the model. The cumulative ability distribution function  $F$  is twice differentiable and its probability density function  $f$  is strictly positive everywhere ( $f > 0$ ). In addition, we impose the assumption that the hazard rate,  $h(\theta) = f(\theta)/(1 - F(\theta))$ , monotonically increases in ability, i.e.,  $h'(\theta) \geq 0$ . The student’s effort cost function  $C$  is separable in test score  $x$  and ability  $\theta$  and takes the form of  $C(x, \theta) = y(x)/\theta$ , where  $y$  is an increasing and strictly convex function of  $x$ . Then, denote the teacher’s utility of test score  $x$  elicited from the student by function  $V : [0, \bar{x}] \rightarrow \mathbb{R}^+$ ,  $V_x > 0$ , and assume, until further notice, that it is linear in  $x$  :

$$V(x) = x. \tag{1}$$

The teacher’s linear utility can be interpreted that she equally cares about the performance of low- and high-ability students. Later in the text, when discussing policy applications of the model, the linear case serves as a benchmark for the cases when 1) function  $V$  is convex (the teacher puts more weight on high-ability students), and 2)

function  $V$  is concave (the teacher puts more weight on low-ability students).

Next, when formulating the teacher’s utility maximization problem—designing the student-knowledge-maximizing grading rule—we, without loss of generality, restrict the set of grading rules to direct grading rules  $\mathbf{m} = \{\mathbf{x}, \mathbf{r}\}$ , where test score schedule  $\mathbf{x} : \Theta \rightarrow [0, \bar{x}]$  and grade schedule  $\mathbf{r} : \Theta \rightarrow [0, \bar{r}]$  impose on the student the truthful revelation of own ability type. Furthermore, we impose that functions  $\mathbf{x}$  and  $\mathbf{r}$  belong to the class of piecewise continuously differentiable functions (piecewise  $\mathcal{C}^1$ ), which also need to be non-decreasing for a grading rule  $\{\mathbf{x}, \mathbf{r}\}$  to be implementable. Given these restrictions on  $\mathbf{x}$  and  $\mathbf{r}$ , denote the families of these functions by  $\mathcal{C}_{\mathbf{x}}^1$  and  $\mathcal{C}_{\mathbf{r}}^1$ , respectively.

### 3 Nominal Value of Grades

In this section, we model that the student values grades at their face value and derives a higher utility from a higher grade independently of the grading rule applied by the teacher. Given a grading rule  $\{\mathbf{x}, \mathbf{r}\}$ , the student’s net utility of reporting a type  $\hat{\theta}$  is equal to

$$U(\theta, \hat{\theta}) = \mathbf{r}(\hat{\theta}) - C(\mathbf{x}(\hat{\theta}), \theta), \quad (2)$$

where parameter  $\theta$  is the student’s ability type. The student’s reservation utility of participating in the class is normalized to zero (for all ability types).

#### 3.1 The Teacher’s Problem

The teacher maximizes with respect to a grading rule  $\{\mathbf{x}, \mathbf{r}\} \in \mathcal{C}_{\mathbf{x}}^1 \times \mathcal{C}_{\mathbf{r}}^1$  her expected utility

$$\int_{\underline{\theta}}^{\bar{\theta}} \mathbf{x}(\theta) dF(\theta) \quad (3)$$

subject to

$$U(\theta, \theta) \geq U(\theta, \hat{\theta}), \quad (4)$$

$$U(\theta, \theta) \geq 0, \quad (5)$$

$$0 < \mathbf{r}(\theta) \leq \bar{r}, \text{ for all } \theta \text{ and } \hat{\theta} \text{ in } \Theta. \quad (6)$$

In the above, (4) is the student’s incentive compatibility constraint, (5) — the individual rationality constraint, where the function  $U$  is defined in (2); the last constraint imposes an upper bound on the teacher’s rewards (though already imposed by requiring  $\mathbf{r} \in \mathcal{C}_{\mathbf{r}}^1$ ).

Under this specification, the model resembles a standard static principal-agent model with hidden information except for the transfer structure. The distinct feature of this model is that, unlike in most agency models, the transfer function  $\mathbf{r}$  does not enter the



principal's (teacher's) utility function but only the agent's (student's), meaning that the principal does not “pay” to motivate the agent. In other words, rewards are costless to give for the principal but of a value to the agent.

Therefore, unlike in models with monetary transfers, in this model with costless rewards we do not have the intercomparison of the agent's and principal's utilities. To solve the model, we approach it differently from the standard solution method attributable to Mirrlees (1971), which main idea is to obtain a functional equation with one unknown by merging the agent's and principal's optimization problems through the transfer function. The key element in solving our model is the observation that it must be optimal for the principal to reward the highest grade of  $\bar{r}$  to the most efficient agent, i.e., in the solution  $\mathbf{r}(\bar{\theta}) = \bar{r}$ , because the reward is costless. This observation allows us to reduce all the constraints (4)–(6) into a single constraint and solve the model using the standard Lagrangean methods.

### 3.2 Solution

We present the solution, the grading rule  $\{\mathbf{x}, \mathbf{r}\}$  maximizing (3) subject to (4)–(6), in Proposition 1 below, relegating the details of solving the model to the Appendix. The main property of the solution is the pooling of ability types from a non-empty interval  $[\theta^*, \bar{\theta}]$  for the highest score-grade allocation, which we discuss more thoroughly later.

**Proposition 1** *The score-grade allocations  $(\mathbf{x}(\theta), \mathbf{r}(\theta))$ , solving the optimization problem (3)–(6), are characterized*

- for ability types  $\theta$  in  $[\underline{\theta}, \theta^*)$ , where

$$\theta^* = \min\{\theta : (1 - F(\theta))/(\theta f(\theta)) \leq 1, \theta \in [\underline{\theta}, \bar{\theta}]\}, \quad (7)$$

by

$$\mathbf{x}(\theta) = y_x^{-1} \left( \frac{f(\theta)\theta^2 y_x(\mathbf{x}(\theta^*))}{f(\theta^*)\theta^{*2}} \right) \quad (8)$$

for the score allocations  $\mathbf{x}$ , and by

$$\mathbf{r}(\theta) = C(\mathbf{x}(\theta), \theta) - \int_{\underline{\theta}}^{\theta} C_{\theta}(\mathbf{x}(\tilde{\theta}), \tilde{\theta}) d\tilde{\theta} \quad (9)$$

for the grade allocations  $\mathbf{r}$ ;

- for ability types  $\theta$  in  $[\theta^*, \bar{\theta}]$  by the grade  $\mathbf{r}(\theta) = \bar{r}$  and the score allocation  $\mathbf{x}(\theta) = \mathbf{x}(\theta^*)$ , where  $\mathbf{x}(\theta^*)$  is found from

$$\bar{r} - C(\mathbf{x}(\theta^*), \theta^*) + \int_{\underline{\theta}}^{\theta^*} C_{\theta}(\mathbf{x}(\theta), \theta) d\theta = 0. \quad (10)$$

**Proof.** See Appendix A.<sup>10</sup> ■

### 3.3 Results

Below we summarize the main properties of the optimal grading rule with nominal grades, presented in Proposition 1.

#### 3.3.1 Pooling at the top

To make it general, one of the main properties of the model studied is the optimality of uniform allocations among most efficient agent types when the principal does not bear or does not internalize the cost of rewarding the agent. This result is in sharp contrast to the “no distortion at the top” property of optimal contracts in most agency models with costly rewards and hidden information.<sup>11</sup>

**Result 1** *In a single-agent and hidden-information agency problem with costless rewards, the principal pools some of the most efficient agent types for a uniform allocation.*

The proof of this result has been given when deriving the condition for the starting point of the pooling interval (7) in Proposition 1 (see Appendix A), where we showed that a uniform score-grade allocation  $(\mathbf{x}(\theta^*), \bar{r})$  applies to all the agent types from the non-empty interval  $[\theta^*, \bar{\theta}]$  (the starting point  $\theta^*$  is bound to be strictly less than the highest ability type  $\bar{\theta}$ , since  $(1 - F(\bar{\theta})) / (\bar{\theta} f(\bar{\theta})) = 0 < 1$ ). Moreover, neither the existence of an upper bound on rewards nor its size, as imposed by constraint (6), is central to the result, what is crucial is the costlessness of rewards. (In a principal-agent model with costly rewards, imposing an upper bound on the reward function does not lead to pooling among the most efficient agent types as long as this constraint is not binding, i.e., when the upper bound is large enough.)

The finding that there is no perfect screening among the most efficient agent types should not be surprising. Suppose it were the case that in the solution only the most efficient type received the highest reward. To make this allocation incentive compatible, the principal would need to suppress the motivation of other types in order to refrain the most efficient type from misreporting. But, as an alternative to the perfect screening, consider the principal marginally “tilting up” the schedule of all the allocations but

---

<sup>10</sup>The problem is solved for the case with the linear utility function  $V$ . However, this has no effect on the qualitative properties of the solution obtained, which are invariant to the form of the teacher’s utility function  $V$  (which needs to be “less convex” than the effort cost function  $C$ ). In particular, the starting point of the pooling interval  $\theta^*$  remains the same for any functional form of  $V$ .

<sup>11</sup>While the “no distortion at the top” property is characteristic of principal-agent models with monetary rewards, see, e.g., Mirrlees (1971), it has also been shown to hold for agency problems with status incentives, see Moldovanu *et al.* (2007). (Recently, there have also been papers in which this property does not hold in the optimum, see Levin (2003) or MacLeod (2003), where the result hinges on the assumption that the agent’s effort is not verifiable unlike in our model.)

the last one—done at no cost—against the corresponding decrease in the performance allocation of the most efficient agent type. This change in the allocations is sure to be expected performance increasing because the gain from it—the increases in performance levels for almost all types—outweighs its corresponding loss—the decrease in performance level of the most efficient type happening with a very small probability. As a result, the principal finds it optimal to increase the probability mass of agent types subject to the highest reward until the gains and losses described offset each other.

Referring back to the teacher-student relationship, where lenient grading is a widespread phenomenon, pooling most efficient types is even more prevalent when the distribution for abilities is more skewed to the end of low types as discussed next.

### 3.3.2 Mismatch between grades and abilities

Here, we establish a relationship between the optimal grading rule and student ability distribution, which has a strong empirical support from the literature on educational measurement, described later in the paper.

Consider two classes of students, who come from two different student populations, where abilities are distributed on the same support  $\Theta$  according to distributions  $F_1$  and  $F_2$ , respectively. Denote the student types from the two classes by  $\theta_1$  and  $\theta_2$ , respectively, and let the student type  $\theta_2$  be smaller than  $\theta_1$  in the likelihood ratio order, i.e.,

$$\frac{f_2(\theta)}{f_1(\theta)} \text{ decreases for all } \theta \text{ in } \Theta,$$

where  $f_1$  and  $f_2$  are the probability density functions of the corresponding distributions. The interpretation of this stochastic dominance condition is that students from the first class are held more able than those from the second class. (Formally, this condition implies that  $\int_{\theta'}^{\theta''} \theta f_1(\theta) d\theta \geq \int_{\theta'}^{\theta''} \theta f_2(\theta) d\theta$  for any interval  $[\theta', \theta''] \subset \Theta$ , or for any restriction of the ability space the expected student ability in the first class is greater than that in the second class.)

Let  $\{\mathbf{x}_1, \mathbf{r}_1\}$  and  $\{\mathbf{x}_2, \mathbf{r}_2\}$  be the solutions to the optimization problem (3)–(6) for the two classes, respectively. Then, the following holds.

**Result 2** *If the student type  $\theta_2$  is smaller than the student type  $\theta_1$  in the likelihood ratio order, then the optimal grade allocations in the two classes satisfy  $\mathbf{r}_2(\theta) \geq \mathbf{r}_1(\theta)$  for every student type  $\theta$  in  $\Theta$ .*

**Proof.** See Appendix B. ■

To put it in words, this result says that the lower the expectations the teacher holds about her student abilities, the more lenient she should be when grading. The intuition behind the optimality of more lenient grading rules in less able classes comes from the

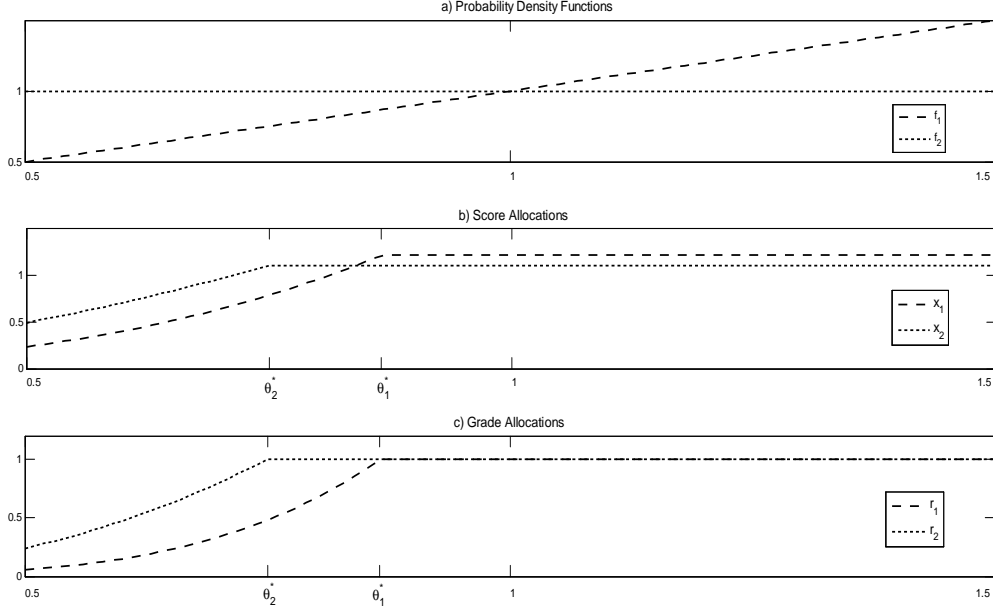


Figure 1: Optimal Grading Rules

teacher’s attempt to extract more effort from more numerous lower-ability student types, and vice versa.

Results 1 and 2 are illustrated in Figure 1. It depicts the optimal grading rules for two classes with the student ability type space  $\Theta = [0.5, 1.5]$ ; the student ability in the first class,  $\theta_1$ , is distributed according to the distribution with  $f_1(\theta) = \theta$ ,  $\theta \in \Theta$ , and in the second class, the student ability  $\theta_2$  is distributed uniformly over  $\Theta$  (the top diagram). The effort cost function takes the form of  $C(x, \theta) = x^2/(2\theta)$  and the upper bound on grades is set to  $\bar{r} = 1$ . The middle diagram of Figure 1 shows the optimal score allocations (the dashed line for the first class and the dotted line for the second class), and the bottom diagram shows the optimal grade allocations for the two classes, respectively. As we can see, both teachers pool some of the most efficient student types. The teacher of the first class, however, offers the highest grade of  $\bar{r}$  to fewer student types but against a higher performance level, while the teacher of the second class optimally chooses to be more lenient. Unlike the teacher of the more able class, the second teacher also offers a flatter score-grade schedule in her attempt to extract more effort from less able but relatively more numerous student types.

## 4 Relative Grades

In this section, we study the situation when the student values a grade only for its value as an ability signal to the job market. In what follows, we distinguish between two cases based on the scope of the market’s knowledge: 1) distinguishable grading rules—the job market observes the grading rule applied to grade the student—and 2) indistinguishable

grading rules—the job market does not observe the exact grading rule applied to grade the student and infers the student’s expected ability from a grade based on its own perception of grading standards applied.

## 4.1 Distinguishable Grading Rules

Here, we modify the previous model by having that 1) the student values a grade for its ability signal, and 2) the job market observes the grading rule set up by the teacher and correctly infers the student’s (expected) ability from his grade, as defined below. Therefore, unlike in the previous section, here the value of a grade is dependent on the distribution of grades across different ability types or, in other words, on the stringency of the grading rule applied by the teacher.

Suppose the teacher designs a grading rule  $\mathbf{m} = \{\mathbf{x}, \mathbf{r}\}$ . Let  $\mathcal{R}(\mathbf{r})$  denote the range of the grade schedule  $\mathbf{r}$ . If  $r \in \mathcal{R}(\mathbf{r})$ ,  $\mathbf{r}^{-1}(r)$  is the set of all  $\theta \in \Theta$  such that  $\mathbf{r}(\theta) = r$ . Let  $\theta_*(r) = \inf(\mathbf{r}^{-1}(r))$  and  $\theta^*(r) = \sup(\mathbf{r}^{-1}(r))$ . Using the monotonicity of grade schedule  $\mathbf{r}$ , we define the ability type  $\theta^{\mathbf{r}}$  inferred by the job market from a grade  $r$ ,  $\theta^{\mathbf{r}} : [0, \bar{r}] \rightarrow \Theta$ , by

$$\theta^{\mathbf{r}}(r) = \begin{cases} \mathbf{r}^{-1}(r) & \text{if } r \in \mathcal{R}(\mathbf{r}) \text{ and } \theta_*(r) = \theta^*(r); \\ \frac{\int_{\theta_*(r)}^{\theta^*(r)} \theta f(\theta) d(\theta)}{F(\theta^*(r)) - F(\theta_*(r))} & \text{if } r \in \mathcal{R}(\mathbf{r}) \text{ and } \theta_*(r) \neq \theta^*(r); \\ 0 & \text{if } r \notin \mathcal{R}(\mathbf{r}). \end{cases} \quad (11)$$

As a technical detail, we set  $\theta^{\mathbf{r}}(r) = 0$  if  $r \notin \mathcal{R}(\mathbf{r})$  to mean that if the student’s grade is outside the range of grades observed then this grade as a signal is meaningless.

The above definition also defines the student’s utility of a grade given the grading rule  $\mathbf{m} = \{\mathbf{x}, \mathbf{r}\}$ , and his net utility of reporting a type  $\hat{\theta}$  is equal to

$$U^{\mathbf{r}}(\theta, \hat{\theta}) = \theta^{\mathbf{r}}(\mathbf{r}(\hat{\theta})) - C(\mathbf{x}(\hat{\theta}), \theta). \quad (12)$$

### The teacher’s problem

As before, the teacher’s problem is to set up a grading rule that elicits from the student the highest expected performance on the test, i.e., the teacher maximizes her expected utility with respect to a grading rule  $\mathbf{m} = \{\mathbf{x}, \mathbf{r}\} \in C_{\mathbf{x}}^1 \times C_{\mathbf{r}}^1$

$$\int_{\underline{\theta}}^{\bar{\theta}} \mathbf{x}(\theta) dF(\theta) \quad (13)$$

subject to

$$U^{\mathbf{r}}(\theta, \theta) \geq U^{\mathbf{r}}(\theta, \hat{\theta}), \quad (14)$$

$$U^{\mathbf{r}}(\theta, \theta) \geq 0, \text{ for all } \theta \text{ and } \hat{\theta} \text{ in } \Theta. \quad (15)$$

In the above, the constraints are the student’s incentive compatibility constraint to report truthfully and individual rationality constraint, respectively, where the student’s utility function  $U^{\mathbf{r}}$  is defined in (12), and  $\theta^{\mathbf{r}}$ , the ability signal for the job market, is defined by (11).

In this version of the model, we can simplify the teacher’s problem (13)–(15) by making it be maximized with respect to a score schedule  $\mathbf{x}$  only. The reason for this is that a grade does not bear any value on its own and, as a result, a score schedule  $\mathbf{x}$  can be implemented by any grade schedule  $\mathbf{r}$  isomorphic to  $\mathbf{x}$  achieving the same utility levels. Therefore, without loss of generality, in the teacher’s problem we restrict the set of grading rules  $\mathcal{C}_{\mathbf{x}}^1 \times \mathcal{C}_{\mathbf{r}}^1$  to such grading rules  $\mathbf{m} = \{\mathbf{x}, \mathbf{r}\}$ , where the grade schedule  $\mathbf{r}$  takes the form of

$$\mathbf{r}(\theta) = \bar{r} \frac{\mathbf{x}(\theta)}{\mathbf{x}(\theta)}, \text{ for any } \theta \in \Theta. \quad (16)$$

All in all, the teacher maximizes (13) with respect to  $\mathbf{x} \in \mathcal{C}_{\mathbf{x}}^1$  subject to (14) and (15) with the grade schedule  $\mathbf{r}$  imposed by (16).

## Solution

If the student values a grade for its ability signal, the teacher, when designing a grading rule, needs to take into account the effect the grading rule itself has on the value of a grade for the student as determined by (11). Unlike in the previous model with nominal grades, it follows that the more lenient the grading rule the teacher sets up, the lower the utility the student gets from a given grade, adversely affecting his learning effort choice decision. Below, we give a condition when the downside effect of leniency in grading—the job market’s “degrading” of students’ grades—outweighs the upside effect, which is more effort extraction from lower ability types.

First, we solve for the optimal score schedule  $\mathbf{x}$  under the conjecture that the teacher perfectly screens all the types.

**Conjecture 1** *The grading rule  $\mathbf{m}$  solving the teacher’s problem (13)–(16) screens every ability type  $\theta$  distinctly.*

As it follows from the conjecture, we need to consider only strictly monotone piecewise  $\mathcal{C}^1$  performance functions  $\mathbf{x}$ . Then, the solution to the teacher’s problem is uniquely characterized by the student’s constraints (14) and (15) only. Given a score schedule  $\mathbf{x}$ , with the grade schedule  $\mathbf{r}$  imposed by (16), the utility to the student of ability  $\theta$  from reporting being a type  $\hat{\theta}$  is given by

$$U^{\mathbf{r}}(\theta, \hat{\theta}) = \hat{\theta} - \frac{y(\mathbf{x}(\hat{\theta}))}{\theta},$$

and, at a point of differentiability, the incentive compatibility constraint implies that

$$\frac{\partial}{\partial \hat{\theta}} U^{\mathbf{x}}(\theta, \theta) = 0,$$

or

$$y_x(\mathbf{x}(\theta))\mathbf{x}_\theta(\theta) = \theta.$$

Next, it has to be that the individual rationality constraint of the lowest ability type  $\underline{\theta}$  needs to be binding (or, more precisely, that of the least efficient type “contracted upon” by the teacher, which we assume to be  $\underline{\theta}$ ). If there are no discontinuities—which we check if it is the case later—the solution to the above differential equation together with the binding individual rationality constraint is

$$\mathbf{x}(\theta) = y^{-1} \left( \frac{\theta^2 + \underline{\theta}^2}{2} \right), \quad (17)$$

for every  $\theta$  in  $\Theta$ . The second-order condition for (17) to be the solution holds. Furthermore, since the derivative of  $\mathbf{x}$  in (17) is positive for every type  $\theta$ , the constraint that  $\mathbf{x}$  be monotone also holds.

But suppose that there are (simple) discontinuities in the score schedule  $\mathbf{x}$  that solve the teacher’s problem. Denote the discontinuity point closest to the type  $\underline{\theta}$  by  $\theta'$ , and let  $\mathbf{x}(\theta' -) \neq \mathbf{x}(\theta')$ , where  $\mathbf{x}(\theta' -)$  is a left-hand limit (the subsequent argument with straightforward alterations also holds for the case  $\mathbf{x}(\theta' +) = \mathbf{x}(\theta')$ , where  $\mathbf{x}(\theta' +)$  is a right-hand limit). Then, (17) holds only for types  $\theta$  in  $[\underline{\theta}, \theta')$ . But the allocation  $\mathbf{x}(\theta')$  cannot be greater than the allocation determined by (17) for  $\theta = \theta'$ , because  $\mathbf{x}(\theta')$  would not be incentive compatible: we could find a type  $\tilde{\theta} \in [\underline{\theta}, \theta')$  such that  $U^{\mathbf{x}}(\theta', \tilde{\theta}) > U^{\mathbf{x}}(\theta', \theta')$  for every  $\hat{\theta} \in (\tilde{\theta}, \theta')$ . Then,  $\mathbf{x}(\theta')$  should be smaller than the allocation determined by (17) for  $\theta = \theta'$ , which, though, is suboptimal from the teacher’s perspective. Hence, there cannot exist a discontinuity at  $\theta'$ , nor at any other point for the same reason.

Therefore, under the conjecture that all types are screened perfectly, (17) uniquely characterizes the optimal score schedule  $\mathbf{x}$  for every  $\theta$  in  $\Theta$ . Next, we need to give a condition when this conjecture is valid. It turns out that the sufficient condition is the convexity of the score schedule  $\mathbf{x}$  in (17).

**Proposition 2** *If the score schedule  $\mathbf{x}$  defined in (17) is convex, then it is optimal for the teacher to screen every student ability type  $\theta \in \Theta$  when solving problem (13)–(16).*

**Proof.** See Appendix C. ■

The idea of the proof is that, once the convexity condition is met, a grading rule containing a uniform allocation for some types can be improved upon by separating those types with distinct (and incentive-compatible) allocations as in (17). Furthermore,

the convexity condition is equivalent to requiring the marginal disutility from effort under  $\mathbf{x}$  as in (17) decrease in ability

$$\frac{\partial}{\partial \theta}(C_x(\mathbf{x}(\theta), \theta)) < 0 \quad (18)$$

or, under the functional assumption of  $C(x, \theta) = y(x)/\theta$ ,

$$y_{xx} \leq \left(\frac{y_x}{\theta}\right)^2,$$

which is to require that the effort cost function be not “too convex” in scores  $x$ . (For instance, this condition is met for the effort cost function quadratic in  $x$  and, correspondingly, for other cost functions “less convex” than the quadratic one).

However, if the score schedule  $\mathbf{x}$  defined in (17) is not convex everywhere, then at the restrictions of the type space  $\Theta$ , where it is concave, we obtain pooling of types (by reversing the argument in the proof of Proposition 2). Moreover, with a general form of the teacher’s utility function  $V$  the condition for screening types is that the function

$$V \left[ y^{-1} \left( \frac{\theta^2 + \underline{\theta}^2}{2} \right) \right]$$

is convex in  $\theta$ . With a concave (convex) utility function  $V$ , this condition holds, correspondingly, less (more) often.

Given that the convexity condition holds, the finding that the teacher screens all the student types—when the student values grades for their signaling value and the job market can observe the grading rule designed—is in stark contrast to the results obtained for the case with nominal grades, where pooling of most efficient types is always optimal (see Result 1 of the previous section). This difference in the optimal grading rules results from the difference in the utility of a grade perceived by the student in the two models studied. However, as we show next, if the job market cannot distinguish between grading rules and relate the student’s grade to the grading rule applied, then we should again observe coarse grading rules with pooling types even if the student values a grade for its signaling value only.

## 4.2 Indistinguishable Grading Rules

Here, we study the situation when the job market does not observe the grading rule  $\mathbf{m}$ —in particular, the grade schedule  $\mathbf{r}$  of  $\mathbf{m}$ —that the teacher has designed to grade her student. (This could happen when there are too many teachers for the job market to distinguish among.) Instead, we assume that the market holds its own perception about the grading standard applied, which we denote by  $\boldsymbol{\rho} \in C_r^1$ , and the student’s expected ability it infers from a grade  $r$  is, accordingly,  $\theta^\rho(r)$  defined by (11) for the grading standard  $\boldsymbol{\rho}$  (i.e., in



definition [11]  $\mathbf{r}$  is replaced with  $\boldsymbol{\rho}$ ). The ability inferred,  $\theta^\rho(r)$ , also defines the student's value of a grade  $r$ . The assumption is that the grading standard  $\boldsymbol{\rho}$  is public information.

#### 4.2.1 The teacher's problem

Now, the teacher's problem is to design a grading rule  $\mathbf{m} = \{\mathbf{x}, \mathbf{r}\} \in C_{\mathbf{x}}^1 \times C_{\mathbf{r}}^1$  such that it maximizes her expected utility against the grading standard  $\boldsymbol{\rho}$

$$\int_{\underline{\theta}}^{\bar{\theta}} \mathbf{x}(\theta) dF(\theta) \quad (19)$$

subject to

$$\theta^\rho(\mathbf{r}(\theta)) - C(\mathbf{x}(\theta), \theta) \geq \theta^\rho(\mathbf{r}(\hat{\theta})) - C(\mathbf{x}(\hat{\theta}), \theta), \quad (20)$$

$$\theta^\rho(\mathbf{r}(\theta)) - C(\mathbf{x}(\theta), \theta) \geq 0, \text{ for all } \theta \text{ and } \hat{\theta} \text{ in } \Theta. \quad (21)$$

Unlike in the previous problem (13)–(15), here the teacher does not internalize the effect her grading rule  $\mathbf{m}$  has on the student's utility of a grade,  $\theta^\rho$ , which is determined by the job market's grading standard  $\boldsymbol{\rho}$ . In this respect, the problem is similar to the one with nominal grades (3)–(6), and, therefore, the solution method, applied in Appendix A when solving the model with nominal grades, would apply here, too.

#### 4.2.2 Consistent grading standard

The focus of the analysis presented below is on the properties of grading standard  $\boldsymbol{\rho}$ , which we require to be consistent with the optimal grade schedule  $\mathbf{r}$  designed by the teacher against the grading standard  $\boldsymbol{\rho}$ . First, we make the following definition.

**Definition 1** *Let a mapping  $\pi : C_{\mathbf{r}}^1 \rightarrow C_{\mathbf{r}}^1$  map a grading standard  $\boldsymbol{\rho}$  into the grade schedule  $\mathbf{r}$  of the grading rule  $\mathbf{m}$  solving the teacher's problem (19)–(21) against the grading standard  $\boldsymbol{\rho}$ .*

For analytical convenience, we assume that for any  $\boldsymbol{\rho}$  there is a unique solution to the teacher's problem and, accordingly, a unique grade schedule  $\mathbf{r}$ . Now, we call a grading standard  $\boldsymbol{\rho}$  consistent if it is a fixed point of the mapping  $\pi$ :

$$\boldsymbol{\rho} = \pi(\boldsymbol{\rho}),$$

meaning that the grade schedule  $\mathbf{r}$  of the teacher's best-response grading rule  $\mathbf{m}$  against the grading standard  $\boldsymbol{\rho}$  is equal to the grading standard  $\boldsymbol{\rho}$  itself.

We can immediately claim the following.

**Proposition 3** *A grading standard  $\rho$  with perfect screening of ability types cannot be consistent.*

This result is a direct consequence of Proposition 1 (and Result 1), where we show that in a model with costless rewards pooling of types is inevitable. In the event that the job market perceives a grading standard  $\rho$  that all student types are screened perfectly, i.e.,  $\rho(\theta_1) > \rho(\theta_2)$  if  $\theta_1 > \theta_2$ ,  $\theta_1, \theta_2 \in \Theta$ , the teacher's problem (19)–(21) becomes essentially identical to that with nominal grades (3)–(6), where the teacher can reward the student with any ability signal  $\theta^\rho$  by manipulating the grade schedule  $\mathbf{r}$ . More precisely, we can make the following identity transformations of (19)–(21) to make it look like (3)–(6): set  $\boldsymbol{\theta}(\theta) \equiv \theta^\rho(\mathbf{r}(\theta))$  and introduce a constraint  $\underline{\theta} \leq \boldsymbol{\theta}(\theta) \leq \bar{\theta}$  for all  $\theta \in \Theta$ . Then, the teacher maximizes her expected utility with respect to  $\{\mathbf{x}, \boldsymbol{\theta}\}$ , and the best-response signal schedule  $\boldsymbol{\theta}$  would pool types  $\theta$  in  $[\theta^*, \bar{\theta}]$ , where  $\theta^*$  defined by (7) in Proposition 1, for the reward of  $\bar{\theta}$ , implying that the grading standard  $\rho$  cannot be consistent.

Next, we show that **there exists a consistent grading standard  $\rho$** —a fixed point of  $\pi$ . Consider a grading standard  $\rho$  such that

$$\rho(\theta) = \begin{cases} r^* & \text{if } \theta \geq \theta^\#, \\ r_* & \text{otherwise.} \end{cases} \quad (22)$$

It says that the job market perceives that student types greater than or equal to some  $\theta^\# \in \Theta$  get the grade  $r^*$ , others—the grade  $r_*$ . Given the step grading standard  $\rho$  in (22), the teacher becomes restricted to designing step grading rules  $\mathbf{m} = \{\mathbf{x}, \mathbf{r}\}$  of the form

$$\mathbf{x}(\theta) = \begin{cases} x^* & \text{if } \theta \geq \theta', \\ x_* & \text{otherwise,} \end{cases} \quad \text{and } \mathbf{r}(\theta) = \begin{cases} r^* & \text{if } \theta \geq \theta', \\ r_* & \text{otherwise,} \end{cases} \quad (23)$$

where the teacher's choice variables are scores  $x^*$ ,  $x_*$  and threshold type  $\theta'$ . The question is if there is a step grading standard  $\rho$  with threshold type  $\theta^\#$  such that the threshold type  $\theta'$  of the step grading rule  $\mathbf{m} = \{\mathbf{x}, \mathbf{r}\}$  solving (19)–(21) is equal to the threshold type  $\theta^\#$  of the grading standard  $\rho$ .

For a given threshold type  $\theta'$ , the optimal levels of scores  $x^*$  and  $x_*$  can be expressed as the continuous functions of  $\theta'$  from the binding *IR* and *IC* constraints, respectively. Therefore, the teacher's expected utility in (19) can be expressed as the continuous function of the threshold types  $\theta'$  and  $\theta^\#$  only, which we denote by  $W : \Theta \times (\underline{\theta}, \bar{\theta}) \rightarrow \mathbb{R}$  (the domain of  $\theta^\#$  is set to be an open set, which later is expanded to  $\Theta$ ). Let

$$\tilde{\theta} = \arg \max_{\theta' \in [\underline{\theta}, \bar{\theta}]} W(\theta', \theta^*) = \omega(\theta^*),$$

which characterize the teacher's best-response threshold type  $\tilde{\theta}$  of the optimal grade schedule  $\mathbf{r}$  as a function of  $\theta^\#$  of the grading standard  $\rho$ . Now, the grading standard

$\rho$  is consistent if  $\tilde{\theta} = \theta^\#$ , i.e., if  $\theta^\#$  is a fixed-point of the solution function  $\omega$  (we have assumed the uniqueness of the solution for every  $\theta^\#$ , which is though irrelevant for the following analysis).

By Berge's Maximum Theorem, the solution function  $\omega : (\underline{\theta}, \bar{\theta}) \rightarrow \Theta$  is continuous (because the function  $W$  is continuous and the constraint correspondence is compact-valued and continuous: for any  $\theta^\#$  the teacher's choice set of  $\theta'$  is the whole ability space  $\Theta$ ). The next step is to apply Brouwer's Fixed Point Theorem, for which we need to expand the domain of  $\omega$  to  $\Theta$ . Define function  $\tilde{\omega} : \Theta \rightarrow \Theta$  by

$$\begin{aligned}\tilde{\omega}(\bar{\theta}) &= \lim_{\theta \rightarrow \bar{\theta}} \omega(\theta), \\ \tilde{\omega}(\theta) &= \omega(\theta), \forall \theta \in (\underline{\theta}, \bar{\theta}), \\ \tilde{\omega}(\underline{\theta}) &= \lim_{\theta \rightarrow \underline{\theta}} \omega(\theta).\end{aligned}$$

The function  $\tilde{\omega}$  is continuous, its domain  $\Theta$  is compact and convex, hence by Brouwer's Fixed Point Theorem it has a fixed point  $\theta^{fp}$

$$\theta^{fp} = \tilde{\omega}(\theta^{fp}),$$

which characterizes the consistent grading rule  $\rho$  with  $\theta^\# = \theta^{fp}$ . Having said that, we establish the following

**Proposition 4** *There exists a consistent grading standard  $\rho$ .*

### 4.2.3 Dynamics of grading standards

Finally, within this modeling framework, we can iteratively analyze the dynamics of grading standards, which we sketchily undertake here under the adaptive expectations paradigm. To make a full-fledged analysis of dynamic properties feasible would require putting more structure on the model; instead, we restrict the analysis only to investigating the direction of the change of the interval of ability types pooled for the highest grade.

Suppose the initial grading standard  $\rho_0 \in C_r^1$  is given and let it be a strictly increasing function of  $\theta$  with  $\rho_0(\bar{\theta}) = \bar{r}$  (i.e., the job market initially perceives that the teacher screens every student type). The resultant grade schedule  $\mathbf{r}_0$ , maximizing the teacher's utility against the grading standard  $\rho_0$ , is given by  $\mathbf{r}_0 = \pi(\rho_0)$ . Suppose the grade schedule  $\mathbf{r}_{i-1}$ ,  $i \geq 1$ , is given. We impose that the job market adjusts its perception of grading standards by adapting the existing grade schedule as its new grading standard, i.e.,  $\rho_i = \mathbf{r}_{i-1}$ ,  $i \geq 1$ .

Consider the grade schedule  $\mathbf{r}_0$ . It has the properties described in Proposition 1: pooling types in  $[\theta^0, \bar{\theta}]$ , where  $\theta^0 = \theta^*$  of Proposition 1, and screening the rest of types. At iteration  $i = 1$ , the grading standard that the job market newly perceives is  $\rho_1 = \mathbf{r}_0$ ,

which takes the form of

$$\rho_1(\theta) = \begin{cases} \bar{r} & \text{if } \theta \geq \theta^0, \\ \mathbf{r}_0(\theta) & \text{otherwise.} \end{cases}$$

Now, the highest reward that the teacher can offer the student against the grading standard  $\rho_1$  is  $\theta^{\rho_1}(\bar{r})$ —where  $\theta^{\rho_1}$  is the (discontinuous) ability signal function defined by (11)—and it has depreciated compared with the highest reward available under  $\rho_0$ .

Next, consider the grade schedule  $\mathbf{r}_1 = \pi(\rho_1)$ . One can show that the starting point  $\theta^1$  of the pooling interval  $[\theta^1, \bar{\theta}]$  of the grade schedule  $\mathbf{r}_1$  is characterized by

$$\frac{1 - F(\theta^1)}{f(\theta^1)} = \theta^1 \frac{y_x(\mathbf{x}(\theta^1))(\mathbf{x}(\theta^1) - \mathbf{x}(\theta'))}{y(\mathbf{x}(\theta^1)) - y(\mathbf{x}(\theta'))},$$

where the score allocation  $\mathbf{x}(\theta^1)$  is designed for the types in the pooled interval  $[\theta^1, \bar{\theta}]$  and  $\mathbf{x}(\theta')$  is the second-highest score allocation. The right-hand side of the above expression is greater than  $\theta^1$ —the second term of the product is greater than 1 (follows from the Mean Value Theorem)—implying that the starting point of the new pooling interval  $\theta^1$  is smaller than  $\theta^0$ . Hence, after the initial iteration the pooling interval expands. Furthermore, with the pooling interval expanded, the range of rewards—ability signals—correspondingly shrinks.

To analyze tractably further change in pooling intervals  $[\theta^i, \bar{\theta}]$ ,  $i \geq 2$ , or to see if the sequence  $\{\theta^i\}$  is convergent or not, we would need to place more structure on the model, which is beyond the scope of this paper. Arguably, with the range of rewards diminished, the teacher is bound to apply more lenient grading rules in her attempt to extract more effort from lower types, for she is not anymore able to provide adequately high incentives for higher types, leading to the gradual depreciation of grading standards in the end.

## 5 Discussion

Here, we discuss the main findings of the models presented here. Significantly, the grading patterns obtained closely match those observed in practice lending credibility to our modeling framework of a teacher-student relationship. In this light, we also argue that our model(s) can be used for policy applications such as designing merit-pay programs for teachers.

### 5.1 Main Findings

#### 5.1.1 Compression of grades and grade inflation

The results obtained here show that if there is a reason to think that the principal does not pay for or internalize the cost of rewarding the agent, then in attempt to elicit on average

more effort the principal chooses to be more generous with rewards than otherwise she would have been. In particular, the “no distortion at the top” property, generally observed in models with costly transfers, does not hold here. Now, referring to our teacher-student relationship studied, if an individual teacher cannot credibly commit to her using the same grading standards the job market holds and the job market cannot distinguish among individual grading rules applied, then we face the situation when the teacher treats grades as costless rewards. As a result, the compression of grades or leniency in grading turns out to be the utility-maximizing outcome of the teacher’s grading rule optimally designed: it aims to extract more effort from lower-ability students with the help of costless good grades. To put it differently, good grades are “the commons” that teachers exploit to their benefit. But as in all problems of the commons, we inevitably obtain the deterioration of the commons, which, in our case, takes the form of grade inflation.

From our model(s) studied, we can distinguish two factors contributing toward grade inflation. First, teachers become more lenient with grading in response to shifts in distribution for student abilities toward the lower end of the ability space (Result 2 of this paper). Second, if an individual teacher finds that her grading practice cannot affect the perception of the job market about the signaling value of grades, then the teacher tends to overuse good grades (for whatever signaling value they carry along).

The two factors can both originate from the same source, namely, the expanding availability of education (see, e.g., McKenzie & Tullock (1981)). Due to an increasing number of educational institutions and study programs in recent decades, a larger number of study places has been offered, resulting in more lower-ability applicants being enrolled (see, e.g., Wilson (1999)). Subsequently, this can lead to the emergence of the first factor discussed. Similarly, with more issuers of educational certificates, the grading standards applied by every issuer or teacher become less identifiable, resulting, correspondingly, in the emergence of the second factor.<sup>12</sup>

### 5.1.2 Mismatch between grades and abilities

Our results obtained, in particular Result 2, can offer an explanation why teachers of classes with less able students are more lenient in grading than others (see, e.g., Goldman & Widawski (1976)). As we argue, this can be an outcome of the optimal design of grading rules—i.e., the expected-performance-maximizing outcome—and not necessarily an outcome of some teachers’ rent-seeking behavior, as sometimes is suggested (e.g.,

---

<sup>12</sup>The expanding availability of education was discussed as a circumstance of the grade inflation phenomenon in McKenzie & Tullock (1981). Their hypothesized link is in the context of the demand for and supply of university openings: in response to a higher competition for students—due to the increasing number of university openings relative to the demand—universities engaged in lowering grading standards in order to attract more students. According to McKenzie & Tullock (1981), this practice eventually led to grade inflation.

Johnson (1997)). In particular, our model (with nominal grades) predicts that in classes with less able students grading rules are (optimally) designed so that high grades are more easily attainable, resulting in a mismatch and low correlation between students' grades and their abilities. For instance, if the population of mathematics students contains more talented people than, say, the population of economics students, then we should observe stricter grading standards applied in mathematics classes (as ample empirical evidence shows to be the case, which is explored in the following section).

Concerning the normative side of the differential grading standards discussed, 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 noting that there it is typically assumed 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, 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 a mechanism astray from the projected goals.

## 5.2 Policy Applications

The modeling framework presented here can, arguably, be used to analyze the implications of the introduction of merit-pay programs for teachers. In recent years, a number of such programs have been introduced in various countries to foster incentives for teachers in their endeavors to motivate more effort from their students (see, Lavy (2002, 2009); Atkinson *et al.* (2004); Lazear (2003)). Typically, these programs offer monetary bonuses to teachers if their students improve upon their previous performance (as measured by their scores achieved on standardized tests). The goals pursued by the developers of such programs—social planners—range from improving average performance (in most cases) to reducing the gap between poor and good performers (as, e.g., in the “No Child Left Behind” initiative in the US).

In terms of our modeling framework, the incentives set forth by merit-pay programs for teachers can have a direct effect on the form of the teacher's utility function,  $V$ , in (1). In the case of the “No Child Left Behind” program, where teachers are rewarded for a reduction in the gap between poor and good performers, the utility function  $V$  could turn concave since the teacher would start putting relatively more weight on the performance of lower-ability students. Then, with a concave utility function  $V$ , the model with nominal grades of Section 3<sup>13</sup> predicts that, compared with the case of linear utility, the gap

---

<sup>13</sup>This model is perhaps more appropriate for modeling grading patterns in high schools, for ability-signaling concerns should be of a lesser magnitude among high-school students.

between low- and high-ability students would diminish. However, this reduction would come from two directions, namely, from the teacher’s demanding more effort from low-ability students and demanding less effort from high-ability students.<sup>14</sup> Hence, according to the model, a negative externality from the introduction of this program can arise: high-ability students can be made get the same grades but for less effort. On the contrary, the gap between poor and good performance increases, if the teacher’s utility function turns convex—as a result, for instance, of the social planner implementing a merit-pay program that rewards for students’ excellence only.

With the help our model(s), we can also offer an insight into the problem whether the university administration should restrict the teachers’ choice of grading rules by imposing relative grading, i.e., grading on a curve. From the perspective of our model with relative grades, when the job market observes the grading rule applied by the teacher, an optimal grading rule is, actually, the one that perfectly screens student types. Then, imposing grading on a curve would have no effect since it were not binding. However, when the job market does not observe grading rules, the teacher faces a commitment problem of not overusing good grades, as we discussed before. In this event, grading on a curve would actually bind and could possibly fix the commitment problem of the teacher. But then, the question is what goals the university administration pursues. If they coincide with the teacher’s, i.e., maximizing student knowledge, then grading on a curve would be a desirable policy. But if the administration aims to maximize the expected wage of its students, assuming it is proportional to the ability signal inferred by the market, then the administration may want, as argued in Chan *et al.* (2007) and Ostrovsky & Schwarz (Forthcoming), to refrain from imposing grading on a curve and rather have grades compressed in order to disclose information about student abilities only coarsely.<sup>15</sup>

---

<sup>14</sup>With the concave utility function  $V$ , the optimal score allocations  $\tilde{\mathbf{x}}(\theta)$  for  $\theta$  in  $[\underline{\theta}, \theta^*)$  are equal to

$$\tilde{\mathbf{x}}(\theta) = y_x^{-1} \left( \frac{f(\theta)\theta^2 V_x [\tilde{\mathbf{x}}(\theta)] y_x [\tilde{\mathbf{x}}(\theta^*)]}{f(\theta^*)\theta^{*2} V_x [\tilde{\mathbf{x}}(\theta^*)]} \right).$$

From this, we can immediately observe that it has to be the case that  $\tilde{\mathbf{x}}(\theta^*) < \mathbf{x}(\theta^*)$ , where  $\mathbf{x}(\theta^*)$  is the pooling-interval effort level of the linear-utility case. This is so because the ratio  $V_x [\tilde{\mathbf{x}}(\theta)] / V_x [\tilde{\mathbf{x}}(\theta^*)]$  is strictly greater than 1, and if  $\tilde{\mathbf{x}}(\theta^*) \geq \mathbf{x}(\theta^*)$  then the whole score allocation schedule  $\tilde{\mathbf{x}}$  is above the allocation schedule  $\mathbf{x}$  of the linear case. But it is not possible because  $\tilde{\mathbf{x}}$  would be the solution to the teacher’s problem in the linear case, not  $\mathbf{x}$ . Then, the ratio  $(V_x [\tilde{\mathbf{x}}(\theta)] y_x [\tilde{\mathbf{x}}(\theta^*)]) / V_x [\tilde{\mathbf{x}}(\theta^*)]$  needs to be greater than  $y_x [\mathbf{x}(\theta^*)]$  at least for some  $\theta$  in  $[\underline{\theta}, \theta^*)$ , otherwise the score allocation of the linear case would do better than  $\tilde{\mathbf{x}}$ . Hence, if for some  $\theta$   $\tilde{\mathbf{x}}(\theta) > \mathbf{x}(\theta)$ , so it is for  $\theta = \underline{\theta}$ . All in all, we have a reduction in the gap between the highest and lowest performances, but this reduction comes from both directions.

<sup>15</sup>For more discussion on and empirical implications of making academic transcripts more informative, see Bar *et al.* (2009).

## 6 Empirical Evidence

Here we present empirical evidence in support of our Result 2—the lower the expectations the teacher holds about her students’ abilities, the more lenient the grading rule she sets up.

In general, to test this theoretical prediction of the model, one would need university data such as student grades and their ability proxy (like their performance on university entry exams or Scholastic Aptitude Test [SAT] scores). Then, roughly speaking, one would compare grading patterns for classes with different student ability distributions and see if the prediction holds. However, 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 the *Journal of Educational Measurement* or *Educational and Psychological Measurement*). Most importantly, those studies without exception report results that are fully in line with the model’s predictions: fields with lower ability students studied as compared with those with higher ability students employ less stringent grading criteria. Even though many of those studies are comprehensive in empirical matters, they lack any rigorous theoretical explanation for this phenomenon. Their explanations mainly hinge on intuition or reference to similar phenomena from the adaptation-level theory in psychological literature. In what follows, we attempt to review in detail some of the empirical studies comparing grading standards over time and in different fields, and to show that our model proves helpful in explaining the empirical evidence observed.

Aiken (1963) is one of the first empirical studies that suggest that 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 for this finding, the study only briefly mentions that it conforms to the adaptation-level theory or central tendency phenomenon, which basically concerns the tendency of supervisors to evaluate the performance of people supervised in relative terms rather than in absolute ones.

A much more comprehensive study Goldman & Widawski (1976) first notes the weaknesses of previous studies on grading patterns because of their using the total grade point average (GPA) as the criterion of grading standards. As they rightly argue, GPAs are not perfectly comparable either over time or among individual students because of the possibly different composition of courses included to compute grade averages. To remedy that, Goldman & Widawski (1976) employ a between-subjects design aimed at making grade 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 the 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 & 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, 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, in giving an explanation for the empirical results obtained, they restrict their argument simply by making a reference to the adaptation-level theory that people are judged in comparison to their peers.

A similar study Goldman & Hewitt (1975), which along with presenting the empirical results (which draw the same conclusions about grading behavior as in the studies mentioned above), also provides a more elaborate theoretical explanation for the results obtained. The authors think 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 & Hewitt (1975) continue, as to be considered an almost inevitable factor in the 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, this personal characteristic, as envisaged by Goldman & Hewitt (1975), is not some intrinsic feature of human behavior but rather the outcome of optimal behavior.

A decade later, Strenta & Elliott (1987) replicated the study of Goldman & 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. Therefore, Strenta & 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. (However, we would argue for the reverse direction of causation: since some fields attract more talented students, professors in those fields will grade their students more strin-

gently, which is optimal in order to extract more effort.) As in previous studies, Strenta & Elliott (1987) argue that these differential grading standards serve to attenuate the correlation between the GPAs and SAT scores of the students, and they also show that the correlation increases sizably if GPAs are adjusted by accounting for differences in departmental grading standards. Finally, a similar study conducted at Duke University (Johnson (2003)) confirmed the conclusions about systematic differences in grading standards from the previous studies.

## 7 Conclusion

In this paper, we consider a teacher-student relationship as a special type of an agency problem featuring costless rewards. Our theoretical predictions offer a good match to grading patterns empirically observed both from the static and dynamic perspectives. This allows us to suggest the chosen framework of a principal-agent model is appropriate to analyze a teacher-student relationship.

## Appendix A. Proof of Proposition 1

Here, we solve the teacher's utility maximization problem (3)–(6), namely, we look for the grading rule  $\{\mathbf{x}, \mathbf{r}\} \in \mathcal{C}_{\mathbf{x}}^1 \times \mathcal{C}_{\mathbf{r}}^1$  that maximizes (3) subject to (4)–(6). To start with, we make the following conjecture.

**Conjecture 2** *In the solution to (3)–(6), all types are perfectly screened.*

To put it differently, we conjecture that the allocation schedules  $\mathbf{x}$  and  $\mathbf{r}$  are strictly increasing in type  $\theta$ .

As it is standard, if  $\{\mathbf{x}, \mathbf{r}\}$  is optimal, then it has to be a boundary solution, where the individual rationality constraint of the lowest type is binding

$$U(\underline{\theta}, \underline{\theta}) = 0,$$

and the utility levels for the rest of types  $\theta \in (\underline{\theta}, \bar{\theta}]$ , as can be expressed from the incentive compatibility constraints, are equal to

$$U(\theta, \theta) = - \int_{\underline{\theta}}^{\theta} C_{\theta}(\mathbf{x}(\tilde{\theta}), \tilde{\theta}) d\tilde{\theta}. \quad (24)$$

Next, we observe that since grades are costless for the teacher to reward, it must be that  $\mathbf{r}(\bar{\theta}) = \bar{r}$ . This observation allows us to obtain from the above expression for the utility

level of the highest type  $\theta = \bar{\theta}$

$$\bar{r} - C(\mathbf{x}(\bar{\theta}), \bar{\theta}) + \int_{\underline{\theta}}^{\bar{\theta}} C_{\theta}(\mathbf{x}(\tilde{\theta}), \tilde{\theta}) d\tilde{\theta} = 0, \quad (25)$$

which eliminates the grade schedule from the maximization problem and combines all the constraints into one (because if it is satisfied, so are all the other constraints by choosing the “right” grade schedule).

Having reduced the initial problem, next we set up the Lagrangean, which is

$$\mathcal{L}(\mathbf{x}, \lambda) = \int_{\underline{\theta}}^{\bar{\theta}} \mathbf{x}(\theta) f(\theta) d(\theta) + \lambda [\bar{r} - C(\mathbf{x}(\bar{\theta}), \bar{\theta}) + \int_{\underline{\theta}}^{\bar{\theta}} C_{\theta}(\mathbf{x}(\tilde{\theta}), \tilde{\theta}) d\tilde{\theta}].$$

The first-order conditions (FOCs) are

$$f(\bar{\theta}) + \lambda [-C_x(\mathbf{x}(\bar{\theta}), \bar{\theta}) + C_{x\theta}(\mathbf{x}(\bar{\theta}), \bar{\theta})] = 0 \quad (26)$$

with respect to allocation  $\mathbf{x}(\bar{\theta})$ , and

$$f(\theta) + \lambda C_{x\theta}(\mathbf{x}(\theta), \theta) = 0. \quad (27)$$

with respect to allocations  $\mathbf{x}(\theta)$ ,  $\theta \in [\underline{\theta}, \bar{\theta}]$ . These FOCs together with reduced constraint (25) should characterize the optimal score schedule  $\mathbf{x}$ .

However, if we combine through the Lagrange multiplier  $\lambda$  the first-order condition (26) with (27), we get

$$\frac{f(\bar{\theta})}{f(\theta)} = \frac{C_x(\mathbf{x}(\bar{\theta}), \bar{\theta}) - C_{x\theta}(\mathbf{x}(\bar{\theta}), \bar{\theta})}{-C_{x\theta}(\mathbf{x}(\theta), \theta)},$$

which should hold for any type  $\theta$ . But if we take the limit  $\theta \rightarrow \bar{\theta}$  on both sides, we find that the left-hand side converges to 1, while the right-hand side — to something strictly greater than one:

$$\frac{C_x(\mathbf{x}(\bar{\theta}), \bar{\theta})}{-C_{x\theta}(\lim_{\theta \rightarrow \bar{\theta}} \mathbf{x}(\theta), \bar{\theta})} + \frac{C_{x\theta}(\mathbf{x}(\bar{\theta}), \bar{\theta})}{C_{x\theta}(\lim_{\theta \rightarrow \bar{\theta}} \mathbf{x}(\theta), \bar{\theta})} \geq \frac{C_x(\mathbf{x}(\bar{\theta}), \bar{\theta})}{-C_{x\theta}(\lim_{\theta \rightarrow \bar{\theta}} \mathbf{x}(\theta), \bar{\theta})} + 1 > 1.$$

Hence, there cannot exist a strictly monotonous score schedule  $\mathbf{x}$  that satisfies the first-order conditions and constraint (25). In other words, there is no shadow price  $\lambda$  that can balance all the incentives when screening all the types. Therefore, Conjecture 2 cannot hold, and, in particular, it is not valid “at the top” of the ability space.

Instead, we proceed by pooling types  $\theta$  which are subject to the uniform allocation with the highest grade of  $\bar{r}$ . Let  $\theta^*$  denote the starting value of the pooling interval  $[\theta^*, \bar{\theta}]$ .

After following the same steps as before, the Lagrangean now takes the form of

$$\mathcal{L}(\mathbf{x}, \lambda) = \int_{\underline{\theta}}^{\theta^*} \mathbf{x}(\theta) f(\theta) d(\theta) + (1 - F(\theta^*)) \mathbf{x}(\theta^*) + \lambda [\bar{r} - C(\mathbf{x}(\theta^*), \theta^*) + \int_{\underline{\theta}}^{\theta^*} C_{\theta}(\mathbf{x}(\tilde{\theta}), \tilde{\theta}) d\tilde{\theta}].$$

The FOCs are

$$f(\theta^*) + (1 - F(\theta^*)) + \lambda(-C_x(\mathbf{x}(\theta^*), \theta^*) + C_{x\theta}(\mathbf{x}(\theta^*), \theta^*)) = 0 \quad (28)$$

with respect to allocation  $\mathbf{x}(\theta^*)$ , and, as before,

$$f(\theta) + \lambda C_{x\theta}(\mathbf{x}(\theta), \theta) = 0 \quad (29)$$

with respect to allocations  $\mathbf{x}(\theta)$ ,  $\theta \in [\underline{\theta}, \theta^*)$ . Combining the two conditions, we get

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

Before deriving the condition for the starting point of the pooling interval  $\theta^*$ , we make the following observation. In the solution, the score schedule  $\mathbf{x}$  needs to be continuous at the starting point of the pooling interval. If it were not, then the reward schedule  $\mathbf{r}$  would also be discontinuous (otherwise, the grading rule would not be incentive compatible). But since grades are costless for the teacher, then at no cost she can improve her utility by tilting up the segment of score allocations to the left from the discontinuity point and accordingly adjusting the grade allocations to meet the incentive compatibility constraints. Hence, the score schedule  $\mathbf{x}$  cannot be discontinuous at the starting point of the pooling interval.

Taking the limit  $\theta \rightarrow \theta^*$  on both sides and using the continuity of  $\mathbf{x}$  at  $\theta^*$  we get the condition for the starting point of the pooling interval  $[\theta^*, \bar{\theta}]$ :

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

or, given our assumption that the effort cost function  $C$  is separable in score and type,  $C(x, \theta) = y(x)/\theta$ , this condition becomes

$$\frac{1 - F(\theta^*)}{f(\theta^*)} = \theta^*. \quad (30)$$

Since there may be no type  $\theta$  in  $[\underline{\theta}, \bar{\theta}]$  for which condition (30) holds<sup>16</sup>, then the starting value of the “pooling at the top” interval is defined as

$$\theta^* = \min\{\theta : (1 - F(\theta))/(f(\theta)\theta) \leq 1, \theta \in \Theta\}, \quad (31)$$

which is (7) in Proposition 1. Note that the expression  $(1 - F(\theta))/(f(\theta)\theta)$  is monotonically decreasing in  $\theta$  due to the monotone hazard rate assumption so that the pooling-interval starting point  $\theta^*$  is uniquely determined.

Suppose that  $\theta^* > \underline{\theta}$ . Denote the score-grade allocation for every type  $\theta$  in  $[\theta^*, \bar{\theta}]$  by  $(\mathbf{x}(\theta^*), \bar{r})$ , where the score allocation  $\mathbf{x}(\theta^*)$  needs still to be determined. From (28) and (30) we express the Lagrange multiplier  $\lambda$  to be equal to

$$\lambda = \frac{f(\theta^*)\theta^{*2}}{y_x(\mathbf{x}^*(\theta^*))}. \quad (32)$$

Plugging the above expression for  $\lambda$  into remaining first-order conditions (29), we get that for every  $\theta$  in  $[\underline{\theta}, \theta^*)$  the optimal score allocation  $\mathbf{x}(\theta)$  is equal to

$$\mathbf{x}(\theta) = y_x^{-1} \left( \frac{f(\theta)\theta^2 y_x(\mathbf{x}(\theta^*))}{f(\theta^*)\theta^{*2}} \right), \quad (33)$$

which is (8) in Proposition 1.

Finally, the highest score allocation,  $\mathbf{x}(\theta^*)$ , can be determined from the constraint

$$\bar{r} - C(\mathbf{x}(\theta^*), \theta^*) + \int_{\underline{\theta}}^{\theta^*} C_\theta(\mathbf{x}(\theta), \theta) d\theta = 0, \quad (34)$$

after plugging in the expression for  $\mathbf{x}(\theta)$  from (33), giving (10) in Proposition 2.

The constraint that the schedule of performance allocations  $\mathbf{x}$  be non-decreasing is met, which follows from (30) and the monotone hazard rate assumption.<sup>17</sup>

The optimal grade allocations  $\mathbf{r}(\theta)$  for  $\theta$  in  $[\underline{\theta}, \theta^*)$  are found from (24) and are equal to

$$\mathbf{r}(\theta) = C(\mathbf{x}(\theta), \theta) - \int_{\underline{\theta}}^{\theta} C_\theta(\mathbf{x}(\tilde{\theta}), \tilde{\theta}) d\tilde{\theta}, \quad (35)$$

which is (9) in Proposition 1, concluding the solution to the optimization problem (3)–(6).

<sup>16</sup>The pooling interval comprises the whole type space if, for instance, student types  $\theta$  are uniformly distributed with  $\underline{\theta} \geq \bar{\theta}/2$ .

<sup>17</sup>From equation (30) it follows that  $f(\theta)\theta/(1 - F(\theta)) < 1$  for  $\theta$  in  $[\underline{\theta}, \theta^*)$ , and from the monotone hazard rate:  $f'(\theta) > -f^2(\theta)/(1 - F(\theta))$ . The two properties ensure that the derivative of (33),  $\partial\mathbf{x}/\partial\theta$ , is positive.

## Appendix B. Proof of Result 2

With reference to the pooling-interval condition (7), define  $g_i(\theta) = (1 - F_i(\theta))/(\theta f_i(\theta))$ , and let  $\theta_i^* = \min\{\theta : g_i(\theta) \leq 1, \theta \in \Theta\}$ ,  $i = 1, 2$ . Since the likelihood ratio order implies the hazard rate order (see Shaked & Shanthikumar (1994)), which is  $f_1(\theta)/(1 - F_1(\theta)) \leq f_2(\theta)/(1 - F_2(\theta))$  for every  $\theta$ , it immediately follows that  $g_1(\theta) \geq g_2(\theta)$ , leading to  $\theta_1^* \geq \theta_2^*$ . Hence, we have  $\mathbf{r}_2(\theta) = \bar{r} \geq \mathbf{r}_1(\theta)$  for  $\theta \in [\theta_2^*, \bar{\theta}]$ .

Next, consider the score allocations  $\mathbf{x}_i(\theta)$ ,  $i = 1, 2$ , for types  $\theta$  in  $[\underline{\theta}, \theta_2^*)$ . Denote the Lagrange multipliers from the two optimization problems by  $\lambda_1$  and  $\lambda_2$ , defined by (32) of Appendix A, respectively. Divide the first-order conditions for  $\mathbf{x}_1$  and  $\mathbf{x}_2$  (29) of Appendix A to obtain for any  $\theta$  in  $[\underline{\theta}, \theta_2^*)$

$$\frac{y_x(\mathbf{x}_2(\theta))}{y_x(\mathbf{x}_1(\theta))} = \frac{\lambda_1 f_2(\theta)}{\lambda_2 f_1(\theta)},$$

which also holds at  $\theta = \theta_2^*$  by the continuity of the score schedule  $\mathbf{x}_2$  at  $\theta = \theta_2^*$  as argued in Appendix A.

Since the highest score  $\mathbf{x}_2(\theta_2^*)$  in the second class must be at least as large as  $\mathbf{x}_1(\theta_2^*)$ , which stems from the second teacher's incentive to expand the pooling interval even further (otherwise the score schedule could be improved upon by tilting it up), then at  $\theta = \theta_2^*$  the left-hand side of the above expression is greater than or equal to 1, and so is the right-hand side. Due to the decreasing likelihood ratio  $f_2/f_1$ , the right-hand side stays greater than 1 for any  $\theta$  in  $[\underline{\theta}, \theta_2^*)$ , and so does the left-hand side, implying that  $\mathbf{x}_2(\theta) \geq \mathbf{x}_1(\theta)$  for every  $\theta$  in  $[\underline{\theta}, \theta_2^*)$ , which subsequently leads to  $\mathbf{r}_2(\theta) \geq \mathbf{r}_1(\theta)$  from (9) in Proposition 1.

## Appendix C. Proof of Proposition 2

Define the teacher's expected utility  $\mathcal{V} : \mathcal{C}_{\mathbf{x}}^1 \rightarrow \mathbb{R}$  from implementing a score schedule  $\mathbf{x}$  by

$$\mathcal{V}(\mathbf{x}) = \int_{\underline{\theta}}^{\bar{\theta}} \mathbf{x}(\theta) f(\theta) d\theta.$$

Let the score schedule  $\mathbf{x}^*$  with the grade schedule  $\mathbf{r}$  imposed by (16) solve the teacher's problem (13)–(15) and suppose that there is a non-empty pooling interval  $[\theta', \theta'']$  at which  $\mathbf{x}^*(\theta) = x'$  and  $\mathbf{r}(\theta) = r'$  (the arguments below are also valid if we consider a half-open or open pooling interval). We need to consider two cases 1)  $\theta' = \underline{\theta}$  and 2)  $\theta' > \underline{\theta}$ , and in each case it is sufficient to restrict attention to those performance allocations that satisfy the optimality conditions: the binding individual rationality constraint of the lowest-ability type and the set of downward-binding incentive compatibility constraints, respectively.

**In the first case**, 1)  $\theta' = \underline{\theta}$ , the score allocation  $\mathbf{x}^*(\theta) = x'$  with grade  $r'$  for all  $\theta$

in  $[\underline{\theta}, \theta'']$  need to satisfy the binding individual rationality constraint of the lowest-ability type:

$$\theta^r(r') - \frac{y(x')}{\underline{\theta}} = 0,$$

or

$$x' = y^{-1}[\theta^r(r')\underline{\theta}],$$

where the function  $y^{-1}$  is the inverse of  $y$  and  $\theta^r$  is the ability type inferred as defined by (11). The teacher's expected utility from implementing the score schedule  $\mathbf{x}^*$  is given by

$$\mathcal{V}(\mathbf{x}^*) = F(\theta'')y^{-1}[\theta^r(r')\underline{\theta}] + \int_{\theta''}^{\bar{\theta}} \mathbf{x}^*(\theta)f(\theta)d\theta.$$

As an alternative to the score schedule  $\mathbf{x}^*$ , consider the following score schedule  $\hat{\mathbf{x}}$  : for  $\theta$  in  $[\underline{\theta}, \theta'']$  set (distinct and incentive compatible) performance allocations  $\hat{\mathbf{x}}(\theta) = y^{-1}[(\theta^2 + \underline{\theta}^2)/2]$ , as in (17), and for  $\theta$  in  $(\theta'', \bar{\theta}]$  —  $\hat{\mathbf{x}}(\theta) = \mathbf{x}^*(\theta)$ . The case of interest is the situation when the monotonicity of the new score schedule  $\hat{\mathbf{x}}$  is preserved. Otherwise, when the monotonicity not preserved, i.e., if  $\hat{\mathbf{x}}(\theta'') > \hat{\mathbf{x}}(\theta) = \mathbf{x}^*(\theta)$  for some  $\theta > \theta''$ , the teacher can increase her expected utility by simply setting  $\hat{\mathbf{x}}(\theta) = \hat{\mathbf{x}}(\theta'')$  for all  $\theta > \theta''$  such that  $\hat{\mathbf{x}}(\theta) < \hat{\mathbf{x}}(\theta'')$  and leave the remaining allocations intact. The teacher's expected utility from implementing the score schedule  $\hat{\mathbf{x}}$  is equal to

$$\mathcal{V}(\hat{\mathbf{x}}) = \int_{\underline{\theta}}^{\theta''} \hat{\mathbf{x}}(\theta)f(\theta)d\theta + \int_{\theta''}^{\bar{\theta}} \mathbf{x}^*(\theta)f(\theta)d\theta.$$

The second terms of  $\mathcal{V}(\mathbf{x}^*)$  and  $\mathcal{V}(\hat{\mathbf{x}})$  are identical, and so any difference in the utilities needs to come from the difference in the first terms. Since the suggested performance allocation schedule  $\hat{\mathbf{x}}$  is convex on the restriction  $[\underline{\theta}, \theta'']$ —from the condition of Proposition 2—then by Jensen's inequality

$$\begin{aligned} \int_{\underline{\theta}}^{\theta''} \hat{\mathbf{x}}(\theta)f(\theta)d(\theta) &\geq F(\theta'')\hat{\mathbf{x}}\left(\int_{\underline{\theta}}^{\theta''} \theta f(\theta)d(\theta)/F(\theta'')\right) = F(\theta'')\hat{\mathbf{x}}(\theta^r(r')) \\ &= F(\theta'')y^{-1}[(\theta^r(r')^2 + \underline{\theta}^2)/2] > F(\theta'')y^{-1}(\theta^r(r')\underline{\theta}), \end{aligned}$$

which is equal to the first term of  $\mathcal{V}(\mathbf{x}^*)$ , and where the last inequality stems from the fact that the arithmetic average is greater than the geometric one and  $y^{-1}$  is strictly increasing.

Hence, instead of pooling ability types at the bottom the teacher can do better by screening them since  $\mathcal{V}(\hat{\mathbf{x}}) > \mathcal{V}(\mathbf{x}^*)$ .

**In the second case**,  $\theta' > \underline{\theta}$ , the allocation  $\mathbf{x}^*(\theta) = x'$  for all  $\theta$  in  $[\theta', \theta'']$  together with grade  $r'$  need to satisfy the downward binding incentive compatibility constraint of

the ability type  $\theta'$ , which can be expressed as

$$\theta^{\mathbf{r}}(r') - \frac{y(x')}{\theta'} = \theta^{\mathbf{r}}(r'') - \frac{y(x'')}{\theta'},$$

where the allocation  $(x'', r'')$  is the best alternative to type  $\theta'$ . Since the teacher screens the types in  $[\underline{\theta}, \theta')$  and the optimal way of doing it is as in (17)—otherwise, the score schedule  $\mathbf{x}^*$  is not optimal—then the score allocation  $x'' = \sup_{\theta < \theta'} \mathbf{x}^*(\theta) = y^{-1}[(\theta'^2 + \underline{\theta}^2)/2]$  and the ability signal  $\theta^{\mathbf{r}}(r'') = \sup_{\theta < \theta'} \theta = \theta'$ . The score allocation  $x'$  can, accordingly, be expressed as

$$x' = y^{-1} \left[ \theta' \theta^{\mathbf{r}}(r') - \frac{\theta'^2 - \underline{\theta}^2}{2} \right],$$

and the resulting expected utility to the teacher from implementing the grading rule  $\mathbf{x}^*$  is equal to

$$\mathcal{V}(\mathbf{x}^*) = \int_{\underline{\theta}}^{\theta'} \mathbf{x}^*(\theta) f(\theta) d\theta + (F(\theta'') - F(\theta')) x' + \int_{\theta'}^{\bar{\theta}} \mathbf{x}^*(\theta) f(\theta) d\theta.$$

Similarly to the previous case, consider the following grading rule  $\hat{\mathbf{x}}$ : for  $\theta$  in  $[\theta', \theta'']$  set performance allocations  $\hat{\mathbf{x}}(\theta) = y^{-1}[(\theta^2 + \underline{\theta}^2)/2]$  and for  $\theta$  in  $[\underline{\theta}, \theta') \cup (\theta'', \bar{\theta}]$  —  $\hat{\mathbf{x}}(\theta) = \mathbf{x}^*(\theta)$ . (The monotonicity of the grading rule  $\hat{\mathbf{x}}$  is preserved on the restriction  $[\underline{\theta}, \theta'']$  by the construction of  $\hat{\mathbf{x}}(\theta)$  for  $\theta$  in  $[\theta', \theta'']$ , and regarding the monotonicity over  $(\theta'', \bar{\theta}]$  the same argument as in the first case studied above applies.) The teacher's expected utility from implementing  $\hat{\mathbf{x}}$  is equal to

$$\mathcal{V}(\hat{\mathbf{x}}) = \int_{\underline{\theta}}^{\theta'} \mathbf{x}^*(\theta) f(\theta) d\theta + \int_{\theta'}^{\theta''} \hat{\mathbf{x}}(\theta) dF(\theta) + \int_{\theta''}^{\bar{\theta}} \mathbf{x}^*(\theta) f(\theta) d\theta.$$

Given the condition that  $\hat{\mathbf{x}}(\theta)$  is convex for  $\theta$  in  $[\theta', \theta'']$ , by Jensen's inequality we have for the second term of  $\mathcal{V}(\hat{\mathbf{x}})$  that

$$\begin{aligned} \int_{\theta'}^{\theta''} \hat{\mathbf{x}}(\theta) dF(\theta) &\geq (F(\theta'') - F(\theta')) \hat{\mathbf{x}}(\theta^{\mathbf{r}}(r')) = \\ &= (F(\theta'') - F(\theta')) y^{-1}[(\theta^{\mathbf{r}}(r')^2 + \underline{\theta}^2)/2] > \\ &> (F(\theta'') - F(\theta')) y^{-1}[\theta' \theta^{\mathbf{r}}(r') - (\theta'^2 - \underline{\theta}^2)/2], \end{aligned}$$

which is the second term of  $\mathcal{V}(\mathbf{x}^*)$ , and where the last inequality stems from the fact that  $\theta^{\mathbf{r}}(r') > \theta'$  and  $y^{-1}$  is strictly increasing. From this derivation, we have again that  $\mathcal{V}(\hat{\mathbf{x}}) > \mathcal{V}(\mathbf{x}^*)$ , which concludes the proof of Proposition 2 that pooling student types cannot be optimal for the teacher.



## References

- Aiken, Jr., Lewis R. 1963. The Grading Behavior of a College Faculty. *Educational and Psychological Measurement*, **23**(2), 319–322.
- Atkinson, Adele, Burgess, Simon, Croxson, Bronwyn, Gregg, Paul, Propper, Carol, Slater, Helen, & Wilson, Deborah. 2004. *Evaluating the Impact of Performance-related Pay for Teachers in England*. CMPO working paper No. 04/113.
- Bar, Talia, Kadiyali, Vrinda, & Zussman, Aaf. 2009. Grade Information and Grade Inflation: The Cornell Experiment. *Journal of Economic Perspectives*, **23**(3), 93–108.
- Bedard, Kelly. 2001. Human Capital versus Signaling Models: University Access and High School Dropouts. *Journal of Political Economy*, **109**(4), 749–775.
- Besley, Timothy, & Ghatak, Maitreesh. 2008. Status Incentives. *American Economic Review*, **98**(2), 206–211.
- Chan, William, Hao, Li, & Suen, Wing. 2007. A Signaling Theory of Grade Inflation. *International Economic Review*, **48**, 1065–1090.
- Dickson, Vaughan A. 1984. An Economic Model of Faculty Grading Practices. *The Journal of Economic Education*, **15**(3), 197–203.
- Dubey, Pradeep, & Geanakoplos, John. 2009. *Grading Exams: 100, 99, 98,... or A, B, C?* Cowles Foundation Discussion Paper No. 1710.
- Goldman, Roy D., & Hewitt, Barbara Newlin. 1975. Adaptation-Level as an Explanation for Differential Standards in College Grading. *Journal of Educational Measurement*, **12**(3), 149–161.
- Goldman, Roy D., & Widawski, Mel H. 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**(2), 355–359.
- Guesnerie, Roger, & Laffont, Jean-Jacques. 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**(3), 329–369.
- Johnson, Valen E. 1997. An Alternative to Traditional GPA for Evaluating Student Performance. *Statistical Science*, **12**(4), 251–269.
- Johnson, Valen E. 2003. *Grade Inflation: A Crisis in College Education*. Springer.

- Kuh, G., & Hu, S. 1999. Unraveling the Complexity of the Increase in College Grades from the Mid-1980s to the Mid-1990s. *Educational Evaluation and Policy Analysis*, **20**, 297–320.
- Lavy, Victor. 2002. Evaluating the Effect of Teachers' Group Performance Incentives on Pupil Achievement. *Journal of Political Economy*, **110**(6), 1286–1317.
- Lavy, Victor. 2009. Performance Pay and Teachers' Effort, Productivity, and Grading Ethics. *American Economic Review*, **99**(5), 1979–2011.
- Lazear, Edward. 2003. Teacher Incentives. *Swedish Economic Policy Review*, **10**, 179–2003.
- Levin, Jonathan. 2003. Relational Incentive Contracts. *American Economic Review*, **93**(3), 835–857.
- MacLeod, W. Bentley. 2003. Optimal Contracting with Subjective Evaluation. *American Economic Review*, **93**(1), 216–240.
- McKenzie, Richard B., & Tullock, Gordon. 1981. *The New World of Economics*. Homewood, Illinois: Richard D. Irwin.
- Mirrlees, James A. 1971. An Exploration in the Theory of Optimum Income Taxation. *Review of Economic Studies*, **38**(2), 175–208.
- Moldovanu, Benny, Sela, Aner, & Shi, Xianwen. 2007. Contests for Status. *Journal of Political Economy*, **115**(2), 338–363.
- Ostrovsky, Michael, & Schwarz, Michael A. Forthcoming. Information Disclosure and Unraveling in Matching Markets. *American Economic Journal: Microeconomics*.
- Sabot, Richard, & Wakeman-Linn, John. 1991. Grade Inflation and Course Choice. *Journal of Economic Perspectives*, **5**, 159–170.
- Shaked, Moshe, & Shanthikumar, J. George. 1994. *Stochastic Orders and Their Applications*. Academic Press, Inc.
- Spence, Michael. 1973. Job Market Signaling. *Quarterly Journal of Economics*, **87**(3), 355–374.
- Strenta, A. Christopher, & Elliott, Rogers. 1987. Differential Grading Standards Revisited. *Journal of Educational Measurement*, **24**(4), 281–291.
- Wilson, Bradford P. 1999. The Phenomenon of Grade Inflation in Higher Education. *National Forum*, **Fall**, 38–41.