A Case for Tiered School Systems∗

Jacopo Bizzotto†  Adrien Vigier‡

August 2022

Abstract

We study a mechanism design problem addressing simultaneously how students should be grouped and graded. We argue that the effort-maximizing school systems exhibit coarse stratification and more lenient grading at the top-tier schools than at the bottom-tier schools. Our study contributes to the ongoing policy debate on school tracking.

JEL classification: D02, D80, I20

Keywords: Education, Moral Hazard, Adverse Selection, Design, Grading

∗This paper supersedes an earlier one entitled “Optimal School Design”. We thank Arjada Bardhi, Antonio Cabrales, Bård Harstad, Nenad Kos, Eirik Kristiansen, Ming Li, Martí Mestieri, Johannes Schneider, Takuro Yamashita, and Andy Zapechelnyuk for useful feedback. The present version benefited greatly from the comments of three anonymous referees.

†Faculty of Social Sciences, OsloMet. Email: jacopo.bizzotto@oslomet.no.

‡School of Economics, University of Nottingham. Email: adrien.vigier@nottingham.ac.uk.
1 Introduction

As students regularly point out, grades can be unfair: two students with identical knowledge often end up with different grades. We should expect grades to be imperfect: teachers make mistakes, students’ performance may be affected by external events, and so on. Since grades are imperfect, which school a student attended ends up conveying valuable additional information. Unsurprisingly, therefore, employers and university admissions officers alike rely not only on grades, but also on the school which a student attended in order to make their decisions.

The previous considerations press us to view the problem of designing an optimal school system as comprising (at least) two equally important parts:

(i) How should students be graded?

(ii) How should students be grouped?

Furthermore, as we argue in this paper, the answers to those questions cannot be disentangled: whether schools should separate students by ability depends on the prevailing grading rules; and whether one grading rule performs better than another hinges on the school considered. We therefore study the dual design problem simultaneously addressing questions (i) and (ii) above. We consider a setting in which individuals differ in ability, and where ability and effort jointly determine the distribution of an individual’s productive value. Ability, effort, and productive value are all unobserved. Each individual chooses a school. Then, at the end of school, every student receives a wage proportional to potential employers’ expectation of his productive value.

In our model, an individual’s school choice thus matters for three reasons: (a) the school that a student went to conveys information regarding the distribution of his productive value; (b) different schools may have different grading rules, making it more or less likely that a student will pass; (c) some schools may be more expensive than others.1

We study the mechanism design problem consisting of choosing an incentive compatible and individually rational school system so as to maximize overall effort, and, therefore, the average productive value in the population.

1In our framework, a student’s peers matter due to their impact on the market’s prior beliefs concerning said student’s productive value, thereby linking our work to the vast literature on statistical discrimination reviewed in Fang and Moro (2011).
Two general principles guide our analysis. Firstly, the best way to motivate effort through grades turns on ability: high-ability students are most efficiently motivated by minimizing the chances of failing high-value individuals; low-ability students are most efficiently motivated by minimizing the chances of passing low-value individuals. The intuition is simple. A high-ability student who exerts effort has high chances of ending up with a large productive value. Therefore, in this case, eliminating false negatives makes it possible to detect shirking. By the same logic, a low-ability student who shirks has low chances of ending up with a large productive value. So eliminating false positives makes it possible in this case to detect effort.

The second guiding principle of our analysis is as follows. As indicated earlier, separating students by ability provides potential employers with information concerning students’ productive values. This, in turn, reduces the importance of the grades. Yet effort is motivated through grades, so separating students tends to lower the effort they exert.

Our paper’s main result combines the above principles to fully solve the problem of the designer. The optimal school system has the following features:

- students whose ability is above a cutoff are separated from those whose ability is below it;
- on each side of the separating cutoff, all students are pooled;
- bottom-tier students are evaluated through a tough grading rule (i.e., generating many false negatives), while top-tier students are evaluated through a lenient grading rule (i.e., generating many false positives).

We go on to show that, at an optimum, students in the top tier exert more effort than those in the bottom tier. By contrast, if ability were observable, the designer could want low-ability students to exert more effort than high-ability ones. Interestingly, this familiar “downward distortion at the bottom” property is not achieved by reducing the informativeness of the grading rule in the bottom-tier schools. Instead, the designer raises the fees of the top-tier schools, thus causing fewer and (on average) more able students to attend those schools. It is the resulting small, elite, environment which in turn induces students in the top tier to exert more effort than those in the bottom tier.

Any optimal mechanism is therefore necessarily such that schools in the top tier are more expensive than those in the bottom tier. One such mechanism achieves budget balance by providing scholarships to students in the bottom tier. We show that if the designer
were forced to charge the same fees in every school, then she could do no better than pooling all students in one school.

Our results make a case in favor of “tracking” (the practice of separating students according to ability). Various forms of tracking exist around the world (Wößmann, 2009). Countries like France and Italy sort students across secondary schools on a voluntary basis; in Germany and Switzerland, students’ achievements in their last years of primary school determine the type of secondary school they will be permitted to attend; in Canada, England, and the U.S., tracking normally takes place within a given school. Proponents of tracking see it as an opportunity to tailor curricula to the students’ needs, and to harness desirable peer effects. Critics blame it for effectively providing students from disadvantaged backgrounds with low-quality education.\(^2\)

We contribute to the debate on tracking by showing that, by allowing incentives to be tailored to ability, tracking may help improve overall student achievement. In particular, we show that tracking can be optimal even without direct peer effects.\(^3\) Two important caveats of our analysis are as follows. Firstly, separating students without at the same time adapting the way students are graded is not just useless, it is in fact detrimental. Secondly, even when it is optimal from the perspective of total welfare, tracking may increase disparities between students of different abilities. A planner more concerned about equity than efficiency may thus find tracking to be undesirable.

The remainder of the paper is organized as follows. The related literature is discussed below. The model is presented in Section 2. Section 3 explains the guiding principles of our analysis, and solves two benchmark problems. The main result of the paper is presented in Section 4: we fully characterize the set of optimal mechanisms. Section 5 examines the pattern of effort induced, and explores several extensions. Section 6 concludes.

\(^2\)Empirical evidence regarding the consequences of tracking is mixed. Duflo, Dupas and Kremer (2011) finds evidence of the positive effects of tracking on student achievement. Other studies document negative effects, including Malamud and Pop-Eleches (2011), Maurin and McNally (2012), and Kerr, Pekkarinen and Uusitalo (2013). Similarly, while Hanushek and Wößmann (2006) concludes that tracking exacerbates disparities, Brunello and Checchi (2007) argues that tracking can reduce disparities. For a comprehensive review of the empirical literature, see Hanushek and Wößmann (2011) and Betts (2011). For a review of the debate in sociology, see Gamoran (2009).

\(^3\)Direct peer effects may take various forms (see e.g. Blume, Brock, Durlauf and Jayaraman (2015) and Tincani (2018)). For example, to the extent that students learn from other students, one’s chances of acquiring a competency might be positively affected by an increase in other students’ ability. Alternatively, students might have homophilous preferences, and learn best when grouped with other students of similar ability to their own.
1.1 Related Literature

Our paper primarily contributes to the vast and diverse economics literature on education, and in particular to the strand of this literature explicitly studying grading. Becker and Rosen (1992) argues that stratifying students into groups in which each has a chance for success may be preferred to the setting of a single national standard. Costrell (1994) puts forth the view that a tough grading rule aggravates inequality, and that an egalitarian planner may therefore prefer a more lenient grading rule; Betts (1998) defends the opposite viewpoint. A signalling theory of grade inflation is developed by Chan, Hao and Suen (2007). Ostrovsky and Schwarz (2010) explores a model of assortative matching and shows that if the assignment function from student ability to job desirability is concave, then a school will find it profitable to mix high-ability students with low-ability ones by lowering the informativeness of the grades. Dubey and Geanakoplos (2010) and Boleslavsky and Cotton (2015) question the relation between the informativeness of the grades and the effort which students exert. None of the aforementioned papers examines the dual design problem simultaneously answering how students should be grouped and graded.

From a theoretical viewpoint, our work belongs to the literature on mechanism design with peer effects, comprising Jehiel, Moldovanu and Stacchetti (1996, 1999), Jehiel and Moldovanu (2001), Board (2009), Rayo (2013), Rothschild and Scheuer (2013, 2016), and Yamashita and Sarkisian (2021). The main novelty of our work is that how much an individual values being part of a certain group is not fixed, but instead depends on a grading rule which the designer can adjust. Moreover, the setting we study features moral hazard in addition to screening; Rothschild and Scheuer (2013, 2016) share this feature with our work, but study optimal taxation problems in which agents must choose between two sectors.

Finally, our paper is related to a small but important strand of literature exploring certification under moral hazard. Albano and Lizzeri (2001), Zubrickas (2015), Zapechelnyuk (2020), and Saeedi and Shourideh (2020) all combine moral hazard and adverse selection; the central question they ask is: how to design ratings so as to incentivize producers to exert the largest possible amount of effort? Boleslavsky and Kim (2018) investigate a somewhat different problem: there is no adverse selection, and the central issue is the tension between being lenient so as to certify the largest possible number of agents, and being tough so as not to dilute the value of certification. Our work distinguishes itself from the aforementioned papers in that potential employers can tap into two sources of information:
the grade which a student obtained, and the school which he attended.\(^4\)

## 2 The Model

We consider an environment in which education gives people an opportunity to increase their productive value, and where effort bolsters an individual’s chances of success. Individuals differ in ability: for a given level of effort, more able individuals have better odds of increasing their value. An individual’s ability, effort and productive value are all unobserved, but schools deliver grades, which provide potential employers with information about a student’s productive value. Schools thereby create incentives to exert effort. We study the mechanism design problem consisting of choosing an incentive compatible and individually rational school system so as to maximize the average productive value in the population. The details of the model are described below.

### 2.1 General Environment

**Productive Value.** The productive value of individual \(j\), written \(v_j\), is drawn from \([0, 1]\) according to

\[
\mathbb{P}(v_j = 1) = \alpha \theta_j + (1 - \alpha) e_j,
\]

where \(\alpha\) is a parameter in \((0, 1)\), \(\theta_j \in [0, 1]\) an exogenous attribute referred to as individual \(j\)’s (ability) type, and \(e_j \in [0, 1]\) a choice variable referred to as individual \(j\)’s effort. Types are distributed according to the atomless distribution function \(F\), with support \([0, 1]\).\(^5\)

Given \(\Theta \subseteq [0, 1]\), we shall refer to \(\int_{\Theta} dF(\theta)\) as the mass of students whose types belong to \(\Theta\).

**Schools.** A school \(s\) consists of a pair \((\Theta_s, G(s))\): the subset \(\Theta_s\) represents the types of the students in school \(s\), and \(G(s)\) said school’s grading rule (described in the next paragraph).

\(^4\)At an abstract level, the problem we analyze can be viewed as an information design problem with constraints: the school system plays the role of an information structure, and incentive compatibility constraints ensure that an information structure is feasible. Information design problems with constraints are explored in many recent papers, including Rosar (2017), Lipnowski, Mathevet and Wei (2020), Bloedel and Segal (2021), Doval and Skreta (2021), Bizzotto, Perez-Richet and Vigier (2021), and Perez-Richet and Skreta (2022).

\(^5\)This inessential assumption simplifies the statements of our results.
For brevity, $m_s$ and $\bar{\theta}_s$ shall respectively denote the mass and average type of the students in school $s$.

**Grading.** A grading rule $G$ is an element from the set

$$G := \{(g_0, \Delta) : g_0 \geq 0, 0 \leq \Delta \leq b, g_0 + \Delta \leq 1\},$$

where $b$ is a fixed parameter. The interpretation is as follows: $G = (g_0, \Delta)$ delivers the grade “pass” to student $j$ with probability $g_0$ (respectively, $g_0 + \Delta$) if $v_j = 0$ (resp., $v_j = 1$), and the grade “fail” otherwise. We say that $G$ is *informative* if $\Delta > 0$. We suppose that

$$0 < b < 1.$$

The first inequality assures that informative grading rules exist; the second one captures the idea that all grading rules are imperfect.

**Students’ Payoffs.** An individual $j$’s ability, effort, and productive value are all private information. It is assumed that potential employers observe only (i) the school which $j$ attended and (ii) the grade which he obtained. The (Bayesian) posterior belief which a potential employer attaches to $v_j = 1$ is written $q_j$. The payoff $\pi_j$ of individual $j$ is then given by

$$\pi_j = q_j - c(e_j, \theta_j) - t(s_j),$$

where $c(e_j, \theta_j)$ represents effort cost and $t(s_j)$ the fee of the school $s_j$ attended by $j$. For tractability, effort costs are quadratic:

$$c(e_j, \theta_j) = \frac{1}{2}ce_j^2 - \kappa \theta_j e_j,$$

where $\kappa \geq 0$ and $c > (1 - \alpha)b + \kappa$.\footnote{The latter restriction will ensure that all students exert an interior level of effort.}

**Consistency of Beliefs.** We show in Appendix A that to each school $s$ corresponds a unique consistent profile of effort, that is, a unique

$$e^*_s : \Theta_s \rightarrow [0, 1]$$
such that, when potential employers believe that students’ effort profile is given by $e^*_s$, then exerting effort $e^*_s(\theta)$ maximizes the expected payoff of a type-$\theta$ individual. For brevity, $\overline{e}_s$ will denote the average consistent effort in school $s$. Then, letting $s_j$ denote the school attended by $j$, potential employers’ prior belief that $v_j = 1$ can be written $p(s_j)$, where

$$p(s) := a\overline{e}_s + (1 - a)\overline{e}_s.$$ 

Hence, $q_j$ equals $q^+(s_j)$ following the grade “pass”, and $q^-(s_j)$ following the grade “fail”, where

$$q^+(s) := \frac{p(s)(g_0(s) + \Delta(s))}{g_0(s) + p(s)\Delta(s)}, \quad q^-(s) := \frac{p(s)(1 - g_0(s) - \Delta(s))}{1 - g_0(s) - p(s)\Delta(s)}.$$ 

### 2.2 The Mechanism Design Problem

**Mechanism.** A school system, generically denoted by $S$, consists of a finite collection of schools, say $\{s_1, \ldots, s_n\}$, such that:

(i) $\Theta_{s_1}, \ldots, \Theta_{s_n}$ form a partition of $[0, 1]$;

(ii) $m_{s_i} > 0$ for all $i$;

(iii) for all $i \neq k$: either $\overline{\theta}_{s_i} \neq \overline{\theta}_{s_k}$ or $G(s_j) \neq G(s_k)$.

A mechanism $M$ consists of a pair $(S, t)$, comprising a school system $S$ and a mapping $t : S \rightarrow \mathbb{R}$ specifying the school fees.\(^8\) \(^9\) \(^10\)

**Timing of Events.** First, the designer proposes a mechanism. Each individual then chooses a school, and how much effort to exert. Next, productive values are drawn and grades

\(^7\)That is, $\overline{e}_s := \frac{1}{m_s} \int_{\Theta_s} e^*_s(\theta)dF(\theta)$.

\(^8\)Our model rules out mechanisms randomizing the allocation of students to schools. As will become clear in the next sections, such mechanisms are never optimal; so ruling them out is inconsequential.

\(^9\)Following the usual interpretation, the designer can be thought of as letting individuals choose from a menu of items. Here, an item consists of a school and a fee. In turn, each school is made up of a collection of students and a grading rule. A number of recent papers share with ours the feature that the designer offers a menu of items comprising a statistical experiment of some kind. This literature includes Esö and Szentes (2007), Kolotilin, Mylovanov, Zapechelnyuk and Li (2017), Li and Shi (2017), Bergemann, Bonatti and Smolin (2018), Guo and Shmaya (2019), Wei and Green (2020), Yamashita and Zhu (2021), and Yang (2022). In all those papers, the experiment helps the agent solve a decision problem. Instead, in our setting the experiment (i.e., the grading rule) helps potential employers evaluate the productive value of the agent (the student).

\(^10\)We allow negative fees to model subsidies (e.g. scholarships, student loans, paid internship programs).
delivered, resulting in a posterior belief $q_j$ regarding student $j$. Finally, every student $j$ receives a payoff of $\pi_j$ given by (1).

**Feasibility.** Let $\psi(\theta,e,s) := q^- + \left[ g_0(s) + \left( \alpha \theta + (1 - \alpha) e \right) \Delta(s) \right] \left[ q^+(s) - q^-(s) \right]$. In words, $\psi(\theta,e,s)$ represents the \textit{gross} expected payoff of a type-$\theta$ student exerting effort $e$ in school $s$. The corresponding expected payoff $U(\theta,e,s)$ of a type-$\theta$ student exerting effort $e$ in school $s$ is given by

$$U(\theta,e,s) = \psi(\theta,e,s) - c(e,\theta) - t(s).$$ (2)

**Definition 1.** A mechanism $M = (S,t)$ is feasible if

$$U(\theta,e^*_s(\theta),s) \geq \max_{s' \in S, e \in [0,1]} U(\theta,e,s'), \text{ for all } s \in S \text{ and } \theta \in \Theta_s,$$

(IC)

and

$$U(\theta,e^*_s(\theta),s) \geq 0, \text{ for all } s \in S \text{ and } \theta \in \Theta_s.$$ (IR)

In words, $M$ is feasible if all individuals whose types belong to $\Theta_s$ prefer choosing $s$ rather than any other school in the system $S$, or dropping out of school and receiving a payoff of 0.

**Designer’s Problem.** The designer’s problem, written $(P_0)$, is to find a feasible mechanism which maximizes overall effort:

$$\max_M \sum_{s \in S} \int_{\Theta_s} e^*_s(\theta) dF(\theta), \text{ s.t. } M \text{ is feasible.}$$ (P_0)

**2.3 Discussion of the Model**

We discuss below several modeling choices we have made:

1. We study the problem of a designer aiming to maximize overall effort (or, equivalently, average productive value). The problem of a designer maximizing total welfare (that is, who internalizes students’ cost of effort) is analyzed in Online Appendix 2: the qualitative features of the optimal mechanisms are identical.

2. The mechanisms we consider allow different schools to have different fees. We show in Subsection 5.2 that with uniform fees the designer can do no better than pooling all students in one school.
3. We show in Subsection 5.3 that requiring every school to use a grading rule which maximizes the effort of its student does not change the set of optimal mechanisms.

4. We show in Appendix D that an optimal mechanism exists which also satisfies budget balance. Ignoring budget balance is thus without loss of generality.

5. We show in Online Appendix 1 that for $\kappa = 0$ all our qualitative results carry through under general convex effort costs. For $\kappa > 0$, as long as effort costs are quadratic the peer effects are pinned down by the mean type of the school considered. Said functional form thus enables us to extend our analysis to cases where effort becomes cheaper as ability increases.

3 Preliminaries

This section contains four subsections. Subsection 3.1 gathers several basic remarks. Subsection 3.2 examines the first guiding principle of our analysis: we show that, depending on their types, different groups of students might be optimally incentivized by different grading rules. Subsection 3.3 examines the second guiding principle of our analysis: we show that if different groups of students are evaluated with the same grading rule, then bunching them together increases the effort they exert. Subsection 3.4 builds on the previous principles in order to solve two benchmark problems of the designer, one in which she can choose any grading rule, but cannot separate students, and one in which all schools are required to have the same (exogenously fixed) grading rule.

3.1 Additional Notation and Basic Remarks

Consider an arbitrary school $s$, and some type $\theta \in \Theta_s$. The effort level $e_s^*(\theta)$ equates the marginal benefit and the marginal cost of effort of a type-$\theta$ student in school $s$, giving, by (2),

$$(1 - \alpha)\Delta(s)(q^+(s) - q^-(s)) = ce_s^*(\theta) - \kappa \theta. \quad (3)$$

Integrating both sides of this equation over $\Theta_s$ and dividing through by $m_s$ yields $(1 - \alpha)\Delta(s)(q^+(s) - q^-(s)) = c\overline{e}_s - \kappa \overline{\theta}_s$, or, equivalently,

$$B(G(s), p(s)) = c\overline{e}_s - \kappa \overline{\theta}_s, \quad (4)$$
where

\[ B(G, p) := (1 - \alpha)\Delta \left( \frac{p(g_0 + \Delta)}{g_0 + p\Delta} - \frac{p(1 - g_0 - \Delta)}{1 - g_0 - p\Delta} \right). \]

Combining (3) and (4), we see that

\[ e^*_s(\theta) = \overline{\epsilon}_s + \frac{\kappa}{c} (\theta - \overline{\theta}_s). \quad (5) \]

We close this subsection by stating three basic results which we will use repeatedly in the rest of the analysis. Henceforth, to shorten notation, define the grading rules

\[ G^* := (0, b), \ G_* := (1 - b, b). \]

Notice that both \( G_* \) and \( G^* \) are maximally informative within the set \( G \) of all grading rules. Furthermore, \( G_* \) avoids false negatives, while \( G^* \) avoids false positives.\(^{11}\)

**Lemma 1.** Suppose \( B(G(s), \alpha \overline{\theta}_s + (1 - \alpha)x) > cx - \kappa \overline{\theta}_s \). Then \( \overline{\epsilon}_s > x \).

**Lemma 2.**

\[
\arg\max_G B(G, p) = \begin{cases} 
\{G^*\} & \text{if } p \in (0, \frac{1}{2}) \\
\{G^*, G_*\} & \text{if } p = \frac{1}{2} \\
\{G_*\} & \text{if } p \in (\frac{1}{2}, 1) 
\end{cases}.
\]

**Lemma 3.** \( p(s) \) increases with \( \overline{\theta}_s \).

The (easy) proofs of these lemmas are in Appendix B. In words, Lemma 1 simply says that any level of effort at which the marginal benefit of effort in a given school is greater than its marginal cost is necessarily less than the consistent level of effort in that school. Lemma 2 shows that, fixing potential employers’ prior belief that a student drawn from \( s \) has productive value 1, the marginal benefit of effort in school \( s \) is maximized either when \( G(s) = G^* \), or when \( G(s) = G_* \). Lemma 3 records the intuitively obvious fact that augmenting the mean type of the students at a given school increases potential employers’ prior belief that a student drawn from said school has productive value 1.

\(^{11}\)That is, \( G_* \) delivers the grade “pass” to any student \( j \) for whom \( v_j = 1 \). Similarly, \( G^* \) delivers the grade “fail” to any student \( j \) for whom \( v_j = 0 \).

\(^{12}\)The qualitative results of our paper would be unchanged if the set \( G \) were instead given by

\[ G := \{(g_0, \Delta) : g_0 \geq \varepsilon, 0 \leq \Delta \leq b, g_0 + \Delta \leq 1 - \varepsilon\}, \]

for some \( \varepsilon > 0 \), that is, if every possible grading rule generated both false positives and false negatives.
3.2 On the Benefits of Separating Students

Our first result shows that the best way to motivate effort through grades turns on ability.

**Proposition 1.** If a school $s$ maximizes the effort of its students, then either (i) $G(s) = G^*$ or (ii) $G(s) = G_*$. Furthermore, there exists a cutoff $\theta^*$ such that (i) holds whenever $\overline{G_s} < \theta^*$ and (ii) holds whenever $\overline{G_s} > \theta^*$.

The first part of Proposition 1 tells us that any school $s$ for which the grading rule $G(s)$ does not belong to $\{G^*, G_*\}$ is such that effort exerted in that school can be increased. This result can be intuitively understood as follows. To start with, notice that any grading rule $G \not\in \{G^*, G_*\}$ can be written as a convex combination of $G^*$, $G_*$, and some uninformative grading rule, $G^\emptyset$ say (see Figure 1 for an illustration). As such, any school $s$ in which $G(s) = G$ may be thought of as covertly randomizing between $G^*$, $G_*$, and $G^\emptyset$. Clearly, moral hazard would be mitigated if, somehow, the grading rule eventually used by $s$ were revealed to potential employers. Yet, students’ marginal benefit of effort in this hypothetical setting can be written as a convex combination of the marginal benefits of effort resulting from the grading rules $G^*$, $G_*$, and $G^\emptyset$, respectively. Students’ marginal benefit of effort being larger in the hypothetical setting than in the true setting, we conclude that at least one of $G^*$, $G_*$, and $G^\emptyset$ induces more effort than $G$. Finally, $G^\emptyset$ being uninformative (and, therefore, generating no effort), either $G^*$ or $G_*$ must induce more effort than $G$.

The second part of Proposition 1 tells us that whether $G^*$ induces more effort than $G_*$
or the other way around depends on the mean type of students in school $s$: $G^*$ maximizes effort whenever $\overline{\theta}_s$ is small, and $G_{\star}$ maximizes effort whenever $\overline{\theta}_s$ is large. The intuitive idea is as follows. First, recall that $G_{\star}$ avoids false negatives. Now consider a school $s$ with $G(s) = G_{\star}$ and a large mean type $\overline{\theta}_s$. A student of $s$ who chooses to exert effort is then very likely to obtain the grade “pass”. So failing reveals that a student probably shirked. In other words, for high $\theta_s$, $G_{\star}$ contains moral hazard by enabling potential employers to detect shirking. Reasoning similarly shows that, for low $\theta_s$, $G^*$ enables potential employers to ascertain that a student exerted effort.

**Proof of Proposition 1:** The first part of the proposition is a simple consequence of Lemmata 1 and 2. We prove below the second part of the proposition. Consider two schools, $s_1$ and $s_2$, sharing the same mean type, with grading rules $G(s_1) = G^*$ and $G(s_2) = G_{\star}$. We claim that $p(s_1) > 1/2$ implies $\overline{e}_{s_2} > \overline{e}_{s_1}$. Suppose $p(s_1) > 1/2$. Then

$$B(G(s_2), \alpha \overline{\theta}_{s_2} + (1 - \alpha) \overline{e}_{s_1}) = B(G(s_2), p(s_1)) > B(G(s_1), p(s_1)) = c \overline{e}_{s_1} - \kappa \overline{\theta}_{s_1} = c \overline{e}_{s_1} - \kappa \overline{\theta}_{s_2}. $$

The first and last equalities follow from $\overline{\theta}_{s_1} = \overline{\theta}_{s_2}$, the inequality is a consequence of Lemma 2 and $p(s_1) > 1/2$, and the penultimate equality is obtained from (4). We thus have $B(G(s_2), \alpha \overline{\theta}_{s_2} + (1 - \alpha) \overline{e}_{s_1}) > c \overline{e}_{s_1} - \kappa \overline{\theta}_{s_2}$ and, by Lemma 1, $\overline{e}_{s_2} > \overline{e}_{s_1}$.

Now, Lemma 3 implies the existence of a cutoff $\theta^*$ such that $p(s) > 1/2$ whenever $\overline{\theta}_s > \theta^*$ and $G(s) = G^*$. Hence, by combining the first part of the proposition with the previous claim, we see that if a school $s$ with $\overline{\theta}_s > \theta^*$ maximizes the effort of its students, then $G(s) = G_{\star}$. One shows analogously that if a school $s$ with $\overline{\theta}_s < \theta^*$ maximizes the effort of its students, then $G(s) = G^*$. 

### 3.3 On the Benefits of Pooling Students

Our second result shows that, given any informative grading rule, separating students into different groups tends to lower their effort.

**Proposition 2.** If two schools share the same informative grading rule, then merging them increases total effort.

---

13Readers familiar with the reputation literature will recognize here a mechanism similar to the one driving the reputation dynamics in Board and Meyer-ter Vehn (2013).
Proof: Consider the binary school system \( S = \{s_1, s_2\} \) such that \( s_1 \) and \( s_2 \) share the same informative grading rule, \( G \) say. Let \( S' = \{s_3\} \), where \( s_3 = ([0,1], G) \). We have to show that

\[
\bar{e}_{s_3} > m_{s_1} \bar{e}_{s_1} + m_{s_2} \bar{e}_{s_2}. \tag{6}
\]

As \( B(G, \cdot) \) is strictly concave and \( \bar{\theta}_{s_3} = m_{s_1} \bar{\theta}_{s_1} + m_{s_2} \bar{\theta}_{s_2} \), using (4) yields

\[
B\left(G, \alpha \bar{\theta}_{s_3} + (1 - \alpha)(m_{s_1} \bar{e}_{s_1} + m_{s_2} \bar{e}_{s_2})\right) > m_{s_1} B(G, p(s_1)) + m_{s_2} B(G, p(s_2))
\]

\[
= m_{s_1} (c \bar{e}_{s_1} - \kappa \bar{\theta}_{s_1}) + m_{s_2} (c \bar{e}_{s_2} - \kappa \bar{\theta}_{s_2})
\]

\[
= c \left( m_{s_1} \bar{e}_{s_1} + m_{s_2} \bar{e}_{s_2} \right) - \kappa \bar{\theta}_{s_3}
\]

Using Lemma 1 now establishes (6). \( \Box \)

The basic intuition behind Proposition 2 is as follows. Separating students according to their types provides potential employers with information about students’ productive values. This additional information (on average) weakens the informative content of the grades, and, consequently, also the benefit of exerting effort. Any rationale for separating students must therefore proceed from the opportunity gained to use different grading rules.

### 3.4 Two Benchmarks

In what follows, let \((P_1)\) denote the auxiliary problem of the designer obtained from \((P_0)\) by adding to it the constraint that \( S \) must comprise a single school. The next result immediately follows from Proposition 1.

**Corollary 1.** There exists a cutoff \( \theta^\dagger \) such that, if the mean type in the population is smaller (respectively, greater) than \( \theta^\dagger \), then the solution of \((P_1)\) entails the grading rule \( G^\star \) (respectively, \( G_{\star} \)).

Next, consider the auxiliary problem \((P_2)\) obtained from \((P_0)\) by adding to it the constraint that all schools in \( S \) have to share the same grading rule. The following corollary immediately follows from Proposition 2.

**Corollary 2.** Any solution of \((P_2)\) comprises a single school.
4 Main Result

In this section, we state and prove the main result of our paper, which characterizes the solution of the designer’s problem \((P_0)\). For brevity, say that a property \(X\) is true for almost all students whose types belong to some \(\Theta \subseteq [0, 1]\), if the mass of students with types in \(\Theta\) for whom \(X\) does not hold is equal to 0.

**Definition 2.** A school system has a two-tier structure if it consists of two schools, and there exists a cutoff \(\hat{\Theta}\) such that almost all students whose types are greater than \(\hat{\Theta}\) attend one school (the “top-tier school”), while almost all students whose types are less than \(\hat{\Theta}\) attend the other school (the “bottom-tier school”).

We can now state our main result.

**Theorem 1.** Any solution of \((P_0)\) comprises a school system which either contains a single school, or exhibits a two-tier structure. In the latter case, the bottom-tier school has grading rule \(G^*\), while the top-tier school has grading rule \(G_*\).\(^{14}\)

Readers uninterested in the technical aspects of this result can skip the rest of this section and jump to Section 5 immediately.

4.1 Proof of Theorem 1

We prove the theorem in two steps. Let \((P_3)\) denote the auxiliary problem of the designer obtained from \((P_0)\) by adding to it the constraint that every \(\Theta_s\) has to be an interval. The following proposition summarizes the first step of our proof of Theorem 1.

**Proposition 3.** Any solution of \((P_3)\) comprises a school system which either contains a single school, or exhibits a two-tier structure. In the latter case, the bottom-tier school has grading rule \(G^*\), while the top-tier school has grading rule \(G_*\).\(^{14}\)

The next proposition summarizes the second step of our proof of Theorem 1.

**Proposition 4.** Any school system \(S\) forming part of a solution of \((P_0)\) is such that, for all \(s \in S\), \(\Theta_s\) is almost an interval.\(^{15}\)

The combination of Propositions 3 and 4 yields Theorem 1. Subsection 4.1.1 proves Proposition 3, and Subsection 4.1.2 proves Proposition 4.

\(^{14}\)In the former case, the grading rule is one of \(G^*\) and \(G_*\).

\(^{15}\)That is, there exists an interval \(\Theta\) for which the symmetric difference \(D\) of \(\Theta_s\) and \(\Theta\) has mass 0.
4.1.1 Step 1 of the Proof of Theorem 1

**Definition 3.** A school system $S$ is regular if for any pair of schools $s_1$ and $s_2$ such that $c \bar{c}_{s_1} - \kappa \bar{\theta}_{s_1} < c \bar{c}_{s_2} - \kappa \bar{\theta}_{s_2}$, we have $\sup \Theta_{s_1} \leq \inf \Theta_{s_2}$.

**Remark 1.** Any regular school system $S$ is characterized by unique cutoffs $\hat{\theta}_0 = 0 < \hat{\theta}_1 < \cdots < \hat{\theta}_{I-1} < 1 = \hat{\theta}_I$ such that to every $s \in S$ is associated a unique index, say $i(s)$, satisfying the following properties:

(i) $\Theta_s \subseteq [\hat{\theta}_{i(s)}-1, \hat{\theta}_{i(s)}]$, 

(ii) $i(s') > i(s) \Leftrightarrow c \bar{c}_{s'} - \kappa \bar{\theta}_{s'} > c \bar{c}_{s} - \kappa \bar{\theta}_{s}$.

**Lemma 4.** Any feasible mechanism comprises a regular school system.

**Proof:** Pick a feasible $M = (S, t)$. Straightforward calculations (see Appendix C) show that there exist $\xi : S \rightarrow \mathbb{R}$ and $\zeta : [0, 1] \rightarrow \mathbb{R}$ such that

$$U(\theta, e_s(\theta), s) = \xi(s) + \zeta(\theta) + \left(\frac{\alpha}{1 - \alpha} + \frac{\kappa}{c}\right) \theta(c \bar{c}_s - \kappa \bar{\theta}_s).$$

(7)

Standard mechanism design arguments complete the proof of the lemma. ■

**Lemma 5.** Any regular school system forms part of a feasible mechanism.

**Lemma 6.** If $G(s) = G_*$, then $c \bar{c}_s - \kappa \bar{\theta}_s$ is an increasing function of $\bar{\theta}_s$.

In the rest of this section, $\theta^\dagger$ will refer to the cutoff defined in Proposition 1.

**Proof of Proposition 3:** Suppose some mechanism $M = (S, t)$ with $|S| \geq 2$ solves $(P_3)$. Let $S = \{s_1, \ldots, s_I\}$, with schools labeled so as to satisfy $\sup \Theta_{s_i} \leq \inf \Theta_{s_k}$ for every $i < k$. Then Lemma 4 gives

$$c \bar{c}_{s_1} - \kappa \bar{\theta}_{s_1} \leq \cdots \leq c \bar{c}_{s_I} - \kappa \bar{\theta}_{s_I}.$$  

(8)

Furthermore, notice that

$$\bar{\theta}_{s_1} > \theta^\dagger,$$  

(9)

for otherwise applying Propositions 1 and 2 would enable us to contradict the optimality of $M$ for $(P_3)$. In turn, the combination of (8)-(9), Propositions 1-2, and Lemma 5 shows,

\footnote{We extend here the definition of $e_s(\theta)$ to all $\theta \in [0, 1]$: for any $\theta \in [0, 1]$, $e_s(\theta)$ represents the optimal level of effort for a type-$\theta$ student attending school $s$ (given by (5)).}
firstly, \( G(s_I) = G_\ast \), and, secondly, \( \bar{\Theta}_{s_i} < \theta^\dagger \) for \( i = 1, \ldots, I-1 \). If not, then bunching together all students attending some school whose mean type is at least as large as \( \theta^\dagger \) so as to construct a new school with grading rule \( G_\ast \) would yield a regular school system generating more effort than the system \( S \) forming part of \( M \).

Next, reason by contradiction and suppose either \( I > 2 \), or \( I = 2 \) but \( G(s_1) \neq G^\ast \). Consider the school system \( S' = \{ s'_1, s'_2 \} \) where

\[
s'_1 = \left( \bigcup_{i=1}^{I-1} \Theta_{s_i}, G^\ast \right), \quad s'_2 = s_I.
\]

By Propositions 1 and 2,

\[
\bar{e}_{s'_1} > \frac{1}{\sum_{i=1}^{I-1} m_{s_i}} \sum_{i=1}^{I-1} m_{s_i} \bar{e}_{s'_i}.
\]

So \( S' \) cannot be regular, for otherwise applying Lemma 5 would yield a contradiction with the optimality of \( M \) for problem \((P_3)\). Hence:

\[
\bar{c}_{s'_1} - \kappa \bar{\Theta}_{s'_1} > \bar{c}_{s'_2} - \kappa \bar{\Theta}_{s'_2}. \tag{10}
\]

Now, given \( 0 \leq z < 1 - \inf \Theta_{s'_2} \), consider the school system \( S''(z) \) obtained from \( S' \) by moving those students of the top-tier school \( s'_2 \) whose types are between \( \inf \Theta_{s'_2} \) and \( \inf \Theta_{s'_2} + z \) into the bottom-tier school \( s'_1 \). That is, \( S''(z) = \{ s''_1(z), s''_2(z) \} \), where

\[
s''_1(z) = \left( [0, \inf \Theta_{s'_2} + z], G^\ast \right), \quad s''_2(z) = \left( (\inf \Theta_{s'_2} + z, 1], G_\ast \right).
\]

The difference

\[
D(z) = \left( \bar{c}_{s''_1(z)} - \kappa \bar{\Theta}_{s''_1(z)} \right) - \left( \bar{c}_{s''_2(z)} - \kappa \bar{\Theta}_{s''_2(z)} \right)
\]

is continuous in \( z \), and, by (10), satisfies \( D(0) > 0 \). So by virtue of the intermediate value theorem, either \( D(z) > 0 \) for all \( z \in (0, 1 - \inf \Theta_{s'_2}) \), or there exists \( \tilde{z} \in (0, 1 - \inf \Theta_{s'_2}) \) such that \( D(\tilde{z}) = 0 \). Suppose that the second case holds (the arguments are similar for the first case,
and therefore omitted). Then
\[
m_{s_1''}(z)(c
e^{x}_i(z) - \kappa \theta_{s_i''}(z)) + m_{s_2''}(z)(c
e^{x}_i(z) - \kappa \theta_{s_i''}(z)) = c\ne^{x}_i(z) - \kappa \theta_{s_i''}(z) \\
> c\ne^{x}_i(0) - \kappa \theta_{s_i}(0) \\
= c\ne^{x}_i - \kappa \theta_{s_i} \\
\geq \sum_{i=1}^{l} m_{s_i}(c\ne^{x}_i - \kappa \theta_{s_i}),
\]
where the first equality follows from \(D(z) = 0\), the strict inequality from Lemma 6, and the weak inequality from (8). Yet,
\[
m_{s_1''}(z)\ne^{x}_i(z) + m_{s_2''}(z)\theta_{s_i''}(z) = \sum_{i=1}^{l} m_{s_i} \theta_{s_i}.
\]
We thus obtain
\[
m_{s_1''}(z)\ne^{x}_i(z) + m_{s_2''}(z)\theta_{s_i''}(z) > \sum_{i=1}^{l} m_{s_i} \theta_{s_i}.
\]

But notice that, by virtue of Lemma 5, \(S''(z)\) forms part of a feasible mechanism. The last inequality therefore contradicts our hypothesis that \(M\) solves \((P_3)\).

4.1.2 Step 2 of the Proof of Theorem 1

Proof of Proposition 4: Let \(S\) be part of a mechanism solving \((P_0)\). By Lemma 4, \(S\) is a regular school system. Let \(\{{\hat{\theta}}_i\}_{i=0}^{l}\) be the set of cutoffs characterizing \(S\) (see Remark 1). Suppose \(l = 1\) (the arguments are similar for \(l > 1\), and therefore omitted), that is,
\[
c\ne^{x}_i - \kappa \theta_{s_i} = c\ne^{x}_k - \kappa \theta_{s_k}, \quad \forall s_i, s_k \in S.
\]
Below, let \(S_< := \{s \in S: \theta_{s} < \theta^\dagger\}\), \(S_> := \{s \in S: \theta_{s} > \theta^\dagger\}\), and \(S_\geq := S \setminus S_<\).

Now reason by contradiction, and suppose that not every schools \(s \in S\) is such that \(\Theta_s\) is almost an interval. We can then find two schools in \(S\), say \(s_1\) and \(s_2\), with \(\theta_{s_1} \leq \theta_{s_2}\), as well as subsets \(\tilde{\Theta}_i \subset \Theta_{s_i}\) for \(i = 1, 2\) satisfying: (a) \(\int_{\tilde{\Theta}_1} dF(\theta) = \int_{\tilde{\Theta}_2} dF(\theta) > 0\), and (b) \(\sup \tilde{\Theta}_2 \leq \inf \tilde{\Theta}_1\). We treat below the case in which \(s_1 \in S_<\) and \(s_2 \in S_>\) (the other cases are analogous, and therefore omitted).
Define $S' = [s'_1,s'_2]$ as follows:

$$s'_1 = \left( \bigcup_{s \in \Theta_s} \Theta_s, G^* \right), \quad s'_2 = \left( \bigcup_{s \in S_2} \Theta_s, G^* \right).$$

Using Propositions 1 and 2 shows

$$\bar{v}'_{s'_1} \geq \frac{1}{m_{s'_1}} \sum_{s \in \Theta_s} m_s \bar{e}_s, \quad \bar{v}'_{s'_2} \geq \frac{1}{m_{s'_2}} \sum_{s \in \Theta_s} m_s \bar{e}_s. \quad (12)$$

Next, define $S'' = [s''_1,s''_2]$ by

$$s''_1 = \left( (\Theta_{s'_1} \setminus \hat{\Theta}_1) \cup \hat{\Theta}_2, G^* \right), \quad s''_2 = \left( (\Theta_{s'_2} \setminus \hat{\Theta}_2) \cup \hat{\Theta}_1, G^* \right).$$

In words, $S''$ is obtained from $S'$ by moving students whose type is in $\hat{\Theta}_1$ from $s'_1$ to $s'_2$, and those whose type is in $\hat{\Theta}_2$ from $s'_2$ to $s'_1$. Note that, by construction,

$$\bar{\theta}_{s'_1} < \bar{\theta}_{s'_1} < \bar{\theta}_{s'_2} < \bar{\theta}_{s''_2}. \quad (13)$$

Next, observe firstly that $B(G^*, p)$ is increasing in $p$ and, secondly, that $B(G^*, p)$ is decreasing in $p$. By Lemma 3, we thus obtain

$$B\left( G(s''_1), p(s''_1) \right) > B\left( G(s'_1), p(s'_1) \right), \quad B\left( G(s''_2), p(s''_2) \right) > B\left( G(s'_2), p(s'_2) \right). \quad (14)$$

In turn,

$$m_{s'_1} \bar{v}_{s'_1} + m_{s'_2} \bar{v}_{s'_2} = m_{s'_1} \cdot \frac{B\left( G(s'_1), p(s'_1) \right) + \kappa \bar{\theta}_{s'_1}}{c} + m_{s'_2} \cdot \frac{B\left( G(s'_2), p(s'_2) \right) + \kappa \bar{\theta}_{s'_2}}{c}$$

$$= \frac{1}{c} \left[ m_{s'_1} B\left( G(s'_1), p(s'_1) \right) + m_{s''_2} B\left( G(s'_2), p(s'_2) \right) + \kappa \left( m_{s'_1} \bar{\theta}_{s''_1} + m_{s''_2} \bar{\theta}_{s''_2} \right) \right]$$

$$< \frac{1}{c} \left[ m_{s'_1} B\left( G(s''_1), p(s''_1) \right) + m_{s''_2} B\left( G(s''_2), p(s''_2) \right) + \kappa \left( m_{s''_1} \bar{\theta}_{s''_1} + m_{s''_2} \bar{\theta}_{s''_2} \right) \right]$$

$$= m_{s''_1} \cdot \frac{B\left( G(s''_1), p(s''_1) \right) + \kappa \bar{\theta}_{s''_1}}{c} + m_{s''_2} \cdot \frac{B\left( G(s''_2), p(s''_2) \right) + \kappa \bar{\theta}_{s''_2}}{c}$$

$$= m_{s''_1} \bar{v}_{s''_1} + m_{s''_2} \bar{v}_{s''_2}.$$

The first and last equalities follow from (4). The second equality is a consequence of $m_{s'} = m_{s''}$, for $i = 1, 2$, and $\sum_{i=1}^2 m_{s'_i} \bar{\theta}_{s'_i} = \sum_{i=1}^2 m_{s''_i} \bar{\theta}_{s''_i}$. The inequality is obtained from (14).
By virtue of Lemma 5, if \( c_{s''_2} - \kappa \overline{\theta}_{s''_2} \) were equal to \( c_{s'_1} - \kappa \overline{\theta}_{s'_1} \) then \( S'' \) would form part of a feasible mechanism. But this cannot be, for \( S \) is part of an optimal mechanism, and we saw above that

\[
\sum_{i=1}^{2} m_{s'_i} e_{s'_i} > \sum_{i=1}^{2} m_{s'_i} e'_{s'_i} \geq \sum_{s \in S} m_{s} e_{s}.
\]

So \( c_{s''_2} - \kappa \overline{\theta}_{s''_2} \neq c_{s'_1} - \kappa \overline{\theta}_{s'_1} \). Say \( c_{s''_2} > c_{s'_1} \) (otherwise, just switch the roles of the two schools in what follows), and let \( S''' = \{s''_1, s''_2\} \), with

\[
s''_1 = (\Theta_{s''_1}, (0, \Delta''')), \quad s''_2 = s'_2,
\]

and \( \Delta''\) chosen so as to guarantee \( c_{s''_1} - \kappa \overline{\theta}_{s''_1} = c_{s'_2} - \kappa \overline{\theta}_{s'_2} \). Then

\[
\sum_{i=1}^{2} m_{s''_i} (c_{s''_i} - \kappa \overline{\theta}_{s''_i}) = c_{s''_2} - \kappa \overline{\theta}_{s''_2}
\]
\[
= c_{s'_2} - \kappa \overline{\theta}_{s'_2}
\]
\[
> c_{s'_2} - \kappa \overline{\theta}_{s'_2}
\]
\[
\geq \frac{1}{m_{s'_2}} \sum_{s \in S} m_{s} (c_{s} - \kappa \overline{\theta}_{s})
\]
\[
= \sum_{s \in S} m_{s} (c_{s} - \kappa \overline{\theta}_{s}).
\]

The strict inequality follows from (13) and Lemma 6. The weak inequality is due to (12), and the last equality is a consequence of (11). As

\[
\sum_{i=1}^{2} m_{s''_i} \overline{\theta}_{s''_i} = \sum_{s \in S} m_{s} \overline{\theta}_{s},
\]

the sequence above yields

\[
\sum_{i=1}^{2} m_{s''_i} e_{s''_i} > \sum_{s \in S} m_{s} e_{s}.
\]

Since, by virtue of Lemma 5, \( S''' \) forms part of a feasible mechanism, the latter inequality contradicts the fact that \( S \) is part of a mechanism solving \((P_0)\).
5 Further Analysis

In this section, we show that the optimal mechanisms exhibit higher fees in the top-tier school than in the bottom-tier one, and generate more effort in the former school than in the latter. We then build on those findings to discuss, firstly, the impact of adverse selection on the optimal school system, and secondly, the problem of the designer with uniform fees. The final subsection explores decentralized grading. All proofs are in Appendix D.

This section’s central result is:

**Theorem 2.** Any solution $M = (S, t)$ of $(P_0)$ comprising a two-tier school system is such that both effort and fees are at least as large in the top-tier school ($s_2$) as in the bottom-tier school ($s_1$): $e_{s_2}^* \geq e_{s_1}^*$ and $t(s_2) > t(s_1)$.

The underlying logic of the theorem is as follows. Consider an optimal mechanism comprising a two-tier school system. Then students in the top-tier school must exert more effort than those in the bottom-tier school (Lemma 4). In turn, as both mean type and effort are larger in the top-tier school than in the bottom-tier school, potential employers’ prior belief that students from the top tier have productive value 1 must be larger than the corresponding belief concerning students from the bottom tier. So students are attracted by the top-tier school, regardless of their types. We conclude that any optimal mechanism comprising a two-tier school system must be such that the fee of the top-tier school is greater than the fee of the bottom-tier school.

**5.1 The Consequences of Adverse Selection**

What is the effect of adverse selection on the optimal school system? We show in Appendix D that neither the number of schools nor the grading rules are affected by adverse selection, only the composition of the bottom and top tiers.

By Lemma 4, incentive compatibility leads the designer to induce students in the top tier to exert more effort than those in the bottom tier. One way of achieving this objective would be to reduce the informativeness of the grading rule in the bottom-tier school, keeping the composition of the two schools fixed. We show, however, that a more efficient way of achieving this objective consists in lowering the fee of the bottom-tier school. The idea is as follows. Lowering the fee of the bottom-tier school induces students originally in the top-tier school to switch to the bottom-tier school. This raises firstly the mean type in
both schools, and secondly (by Lemma 3) also the prior beliefs \( p(s_1) \) and \( p(s_2) \) that students from each tier have productive value 1. Yet \( B(G_*, p) \) increases with \( p \), whereas \( B(G^*, p) \) decreases with \( p \). So incentives to exert effort rise in the top-tier school, while simultaneously falling in the bottom-tier school. Lowering the fee of the bottom-tier school thus achieves incentive compatibility.

### 5.2 Uniform Fees

The problem \((P_0)\) allows different schools to have different fees. In practice, a social planner may find that imposing differentiated fees is politically infeasible. This raises the following question: what is the optimal school system under uniform fees?

The next proposition answers the question above. Let \((P_4)\) denote the auxiliary problem of the designer obtained from \((P_0)\) by adding to it the constraint that school fees have to be uniform.\(^\text{17}\)

**Proposition 5.** Suppose that the mean of \( F \) is different from the cutoff \( \theta^\dagger \) in Proposition 1. Then any solution of \((P_4)\) comprises a single school.

Interestingly, the designer is unable to gain from separating students if not at the same time able to impose different fees. Either the mean of \( F \) is larger than \( \theta^\dagger \), in which case the designer can do no better than pooling all students in one school with grading rule \( G_\ast \), or the mean of \( F \) is less than \( \theta^\dagger \) and then the designer can do no better than pooling all students in one school with grading rule \( G^\ast \).

### 5.3 Decentralized Grading

The problem \((P_0)\) enables the designer to determine schools’ grading rules. In practice, a social planner may find it difficult to control the grading rules actually used by schools. Our next and last result shows that if schools care about effort, then the choice of grading rules can be decentralized.

**Proposition 6.** Consider a school \( s \) forming part of a solution of \((P_0)\). Then \( G(s) \) maximizes the effort of the students in school \( s \).\(^\text{17}\)

---

\(^{17}\)That is, \( t(s) = t(s') \) for all \( s, s' \in S \).
6 Conclusion

Because school grades are imperfect, which school a student attended may carry valuable information on top of the grades which he obtained. The problem of designing an optimal school system is thus made up of two parts: how to grade students, and how to group them into schools.

The optimal school system exhibits a mixture of separation and pooling. High-ability students are separated from low-ability ones, because students with different abilities are optimally incentivized to exert effort by different grading rules. On the other hand, all high-ability students are pooled, and so are all low-ability students. The reason is that pooling minimizes the informational content of the school which a student attended, thus making grades matter more. Since effort is motivated through grades, pooling students with similar abilities results in more effort.

The optimal school system is such that students in the bottom tier are evaluated through a tough grading rule, while those in the top tier are evaluated through a lenient grading rule. At an optimum, students in the top tier exert more effort than those in the bottom tier and effectively subsidize the latter.

Our study contributes to the ongoing policy debate on tracking, by showing that tracking can help optimizing students’ incentives to work hard.
Appendix A: Additional Material for Section 2

Proof that to any school \( s \) is associated a unique consistent profile of effort: The special case in which \( \Delta(s) = 0 \) is trivial. Below, suppose \( \Delta(s) > 0 \). We start by showing that any consistent profile of effort \( e^*_s(\cdot) \) must satisfy (4) and (5).

Pick an arbitrary \( \theta \in \Theta_s \). Maximizing (2) with respect to \( e \) yields

\[
B\left(G(s), p(s)\right) = ce^*_s(\theta) - \kappa\theta. \tag{15}
\]

Integrating both sides of this equation over \( \Theta_s \) and dividing through by \( m_s \) yields (4). To obtain (5), just combine (4) with (15).

Next, consider an effort profile \( e^*_s(\cdot) \) satisfying (4) and (5). Observe that in this case (15) holds as well. The function \( U(\theta, e, s) \) (given by (2)) being strictly concave in \( e \), we conclude that, when potential employers believe that students’ effort profile is given by \( e^*_s(\cdot) \), exerting effort \( e^*_s(\theta) \) maximizes the expected payoff of a type-\( \theta \) individual. So (4) and (5) are necessary and sufficient conditions for an effort profile \( e^*_s(\cdot) \) to be consistent.

We now argue that the aforementioned conditions determine a unique profile of effort. The function \( h_s : [0, 1] \to \mathbb{R} \) given by

\[
h_s(x) := B\left(G(s), \alpha \overline{\theta}_s + (1 - \alpha)x\right) - (cx - \kappa \overline{\theta}_s) \tag{16}
\]

is continuous and strictly concave. As \( \overline{\theta}_s \in (0, 1) \), observe that \( 0 < \alpha \overline{\theta}_s + (1 - \alpha)x < 1 \) for all \( x \in [0, 1] \). So \( h_s(0) > 0 \). On the other hand, the assumption \( c > (1 - \alpha)b + \kappa \) implies \( h_s(1) < 0 \). By virtue of the intermediate value theorem, we conclude that \( h_s(x) = 0 \) admits a solution. Said solution is unique, since \( h_s \) is strictly concave.
Appendix B: Additional Material for Section 3

**Proof of Lemma 1:** We saw in Appendix A that \( \overline{s}_x \) is the unique solution of \( h_s(x) = 0 \), where \( h_s \) is given by (16). Furthermore, we noted that \( h_s \) is continuous, concave, and satisfies \( h_s(0) > 0 \) whenever \( \Delta(s) > 0 \). The lemma immediately follows from these observations. ■

**Proof of Lemma 2:** The following properties are readily verified:

(a) for all \( p \in (0, 1) \), \( B\left((g_0, \Delta), p\right) \) is increasing in \( \Delta \) and, for \( \Delta > 0 \), strictly convex in \( g_0 \);

(b) \( B\left((0, b), p\right) \) is decreasing in \( p \), while \( B\left((1 - b, b), p\right) \) is increasing in \( p \);

(c) \( B\left((0, b), \frac{1}{2}\right) = B\left((1 - b, b), \frac{1}{2}\right) \).

Property (a) shows that \( \arg \max_G B(G, p) \subseteq \{G^*, G_s\} \), for every \( p \in (0, 1) \). Combining properties (b) and (c) finishes to prove what we want. ■

**Proof of Lemma 3:** Consider two schools, \( s_1 \) and \( s_2 \), sharing the grading rule, \( G(s_1) = G(s_2) = G \), but with \( \overline{\theta}_{s_1} < \overline{\theta}_{s_2} \). Let \( \tilde{e}_{s_2} \) be defined by

\[
a \overline{\theta}_{s_2} + (1 - a) \tilde{e}_{s_2} = p(s_1).
\]

Notice that \( \tilde{e}_{s_2} < \overline{s}, \) since \( \overline{\theta}_{s_1} < \overline{\theta}_{s_2} \). We then have

\[
B\left(G, a \overline{\theta}_{s_2} + (1 - a) \tilde{e}_{s_2}\right) = B\left(G, p(s_1)\right) = c\tilde{e}_{s_1} - \kappa \overline{\theta}_{s_1} > c\tilde{e}_{s_2} - \kappa \overline{\theta}_{s_2}.
\]

Thus, by Lemma 1, \( \tilde{e}_{s_2} < \overline{\theta}_{s_2} \). It immediately follows that \( p(s_2) > p(s_1) \). ■
Appendix C: Additional Material for Section 4

Proof of (7): We have

\[
U(\theta, e^*_s(\theta), s) = q^-(s) + \left[ g_0(s) + \left( \alpha \theta + (1 - \alpha) e^*_s(\theta) \right) \Delta(s) \right] (q^+(s) - q^-(s)) - \left( \frac{c}{2} e^*_s(\theta) - \kappa \theta \right) e^*_s(\theta) - t(s)
\]

\[
= q^-(s) + \left( \frac{g_0(s)}{\Delta(s)} + \alpha \theta \right) \left( c e_s - \kappa \bar{\theta}_s \right) + e^*_s(\theta) \left( c e^*_s(\theta) - \kappa \theta \right) - \left( \frac{c}{2} e^*_s(\theta) - \kappa \theta \right) e^*_s(\theta) - t(s)
\]

\[
= q^-(s) + \left( \frac{g_0(s)}{\Delta(s)} + \alpha \theta \right) \left( c e_s - \kappa \bar{\theta}_s \right) + \frac{c}{2} \left( e^*_s(\theta) \right)^2 - t(s)
\]

\[
= q^-(s) + \left( \frac{g_0(s)}{\Delta(s)} + \alpha \theta \right) \left( c e_s - \kappa \bar{\theta}_s \right) + \frac{c}{2} \left( e^*_s(\theta) \right)^2 + \frac{\kappa \theta^2}{2c} + \frac{\kappa \theta}{c} \left( c e_s - \kappa \bar{\theta}_s \right) - t(s)
\]

\[
= \xi(s) + \zeta(\theta) + \left( \frac{\alpha}{1 - \alpha} + \frac{\kappa}{c} \right) \theta \left( c e_s - \kappa \bar{\theta}_s \right),
\]

where

\[
\xi(s) = q^-(s) + \frac{g_0(s) \left( c e_s - \kappa \bar{\theta}_s \right)}{(1 - \alpha) \Delta(s)} + \frac{(c e_s - \kappa \bar{\theta}_s)^2}{2c} - t(s), \quad (17)
\]

and

\[
\zeta(\theta) = \frac{\kappa \theta^2}{2c}. \quad (18)
\]

In the sequence above, the second equality follows from (3), and the fourth from (5). \( \blacksquare \)

Proof of Lemma 5: Pick an arbitrary regular school system \( S \), with cutoffs \( \hat{\theta}_0 = 0 < \hat{\theta}_1 < \cdots < \hat{\theta}_{I-1} < 1 = \hat{\theta}_I \) (see Remark 1). For \( i = 1, \ldots, I \), define \( S_i := \{ s : i(s) = i \} \), and let \( (s_1, \ldots, s_I) \) be an arbitrary tuple in \( S_1 \times \cdots \times S_I \). Set \( t(s_1) = 0 \) and, for \( i \geq 2 \), \( t(s_i) \) in such a way that

\[
U\left( \hat{\theta}_{i-1}, e^*_{s_{i-1}}(\hat{\theta}_{i-1}), s_{i-1} \right) = U\left( \hat{\theta}_{i-1}, e^*_{s_i}(\hat{\theta}_{i-1}), s_i \right). \quad (19)
\]

Lastly, for any \( s'_i \in S_i \setminus \{ s_i \} \), set \( t(s'_i) \) so as to obtain\(^{18}\)

\[
U\left( \theta, e^*_{s_i}(\theta), s'_i \right) = U\left( \theta, e^*_{s_i}(\theta), s_i \right), \quad \forall \theta \in [0, 1]. \quad (20)
\]

Now let \( M \) denote the resulting mechanism. Then (19) and (20) imply (IC). Furthermore,\(^{18}\) By (7), such a \( t(s'_i) \) exists and is unique.
(17) and (18) respectively show \( \xi(s_1) \geq 0 \), and \( \zeta(\theta) \geq 0 \) for all \( \theta \in [0,1] \). Hence, by (7),

\[
U\left( \theta, e^{*}_{s_1}(\theta), s_1 \right) \geq 0, \; \forall \theta \in [0,1].
\] (21)

Coupling (IC) and (21) establishes (IR).

Proof of Lemma 6: Notice to begin with that \( B(G_p,p) \) is increasing in \( p \). Thus, if \( G(s) = G_p \), Lemma 3 tells us that \( B\left(G(s),p(s)\right) \) is increasing in \( \bar{\theta}_s \). Equation (4) completes the proof of the lemma.
Appendix D: Additional Material for Section 5

Proof of Theorem 2: Using Lemma 4 shows that 
\[ c e^*_s - \kappa \bar{\theta}_s \leq c e^*_s - \kappa \bar{\theta}_s. \] 
So 
\[ e^*_s \leq e^*_s - \frac{\kappa}{c} (\bar{\theta}_s - \bar{\theta}_s) \leq e^*_s. \] 
In turn, and since \( \bar{\theta}_s > \bar{\theta}_s \), using the law of iterated expectations (twice) yields 
\[ U(\bar{\theta}_s, e^*_s, s) > a \bar{\theta}_s + (1 - a) e^*_s - c e^*_s + t(s) = U(\bar{\theta}_s, e^*_s, s) + (t(s) - t(s)). \] 
On the other hand, \( \Theta_s \) being convex, the feasibility of \( \beta \) implies 
\[ U(\bar{\theta}_s, e^*_s, s) \geq U(\bar{\theta}_s, e^*_s, s). \] 
Yet, by (5), \( e^*_s(\bar{\theta}_s) = e^*_s \). So combining the last two highlighted expressions gives \( t(s) > t(s) \). 

Proof of Proposition 5: Let \( M = (S, t) \) solve \( (P_4) \). Reason by contradiction, and suppose \( |S| \geq 2 \). Let \( s_1 \) and \( s_2 \) be two different schools forming part of \( S \), with \( \bar{\theta}_{s_1} \leq \bar{\theta}_{s_2} \). The arguments used in the proof of Theorem 2 show that \( \bar{\theta}_{s_1} < \bar{\theta}_{s_2} \Rightarrow t(s_1) < t(s_2) \). Since \( t(s_1) = t(s_2) \) we conclude that \( \bar{\theta}_{s_1} = \bar{\theta}_{s_2} \). Yet, by Lemma 4, 
\[ c e^*_s - \kappa \bar{\theta}_s = c e^*_s - \kappa \bar{\theta}_s, \] 
and so \( e^*_s = e^*_s \). These findings being independent of the schools considered, we see that \( \bar{\theta}_{s_i} = \bar{\theta}_{s_i} \) and \( e^*_i = e^*_i \) for all \( s_i, s_k \in S \). Applying Propositions 1 and 2 then yields a contradiction with the optimality of \( M \) for problem \( (P_4) \) (we use here the hypothesis that the mean of \( F \) is not \( \theta^* \)). 

Proof of Proposition 6: Let \( M = (S, t) \) solve \( (P_0) \). By Theorem 1, either \( |S| = 1 \) or \( S \) has a two-tier structure. The result of the proposition is immediate in the former case; combining Propositions 1 and 2 establishes the result in the latter case. 

Proposition D.1. A solution of \( (P_0) \) exists which satisfies budget balance. 

Proof: The arguments are similar to those used in the proof of Lemma 5. The details are standard, and therefore omitted. 

In the remainder of this appendix, we briefly analyze the problem of the designer when types are observable, that is, when (IC) is not required to hold. Below, let \( (P'_0) \) denote
the auxiliary problem of the designer obtained from \((P_0)\) by removing from it the (IC) constraints.

**Proposition D.2.** Any solution of \((P_0')\) comprises a school system which either contains a single school, or exhibits a two-tier structure. In the latter case, the bottom-tier school has grading rule \(G^*\), while the top-tier school has grading rule \(G_\star\).\(^{19}\)

**Proof of Proposition D.2:** Let \(M = (S, t)\) solve \((P_0')\). It is plain that any school \(s\) forming part of \(S\) must maximize the effort of its students. Hence, by Proposition 1, \(G(s) = G^*\) if \(\bar{\theta}_s < \theta^t\) and \(G(s) = G_\star\) if \(\bar{\theta}_s > \theta^t\). Using Proposition 2 then shows that either \(S = \{([0, 1], G^*)\}\), or \(S = \{([0, 1], G_\star)\}\), or else \(S = \{s_1, s_2\}\), with \(\bar{\theta}_{s_1} < \theta^t < \bar{\theta}_{s_2}\), \(G(s_1) = G^*\), and \(G(s_2) = G_\star\). Suppose that we find ourselves in the third case listed above. Reason by contradiction, and suppose that there exist subsets \(\bar{\Theta}_i \subset \Theta_{s_i}\) for \(i = 1, 2\) satisfying: (a) \(\int_{\bar{\Theta}_1} dF(\theta) = \int_{\bar{\Theta}_2} dF(\theta) > 0\), and (b) sup \(\bar{\Theta}_2 \leq \inf \bar{\Theta}_1\). Consider the school system \(S' = \{s_1', s_2'\}\) where

\[
s'_1 = (\Theta_{s_1} \setminus \bar{\Theta}_1) \cup \bar{\Theta}_2, G^*), \quad s'_2 = (\Theta_{s_2} \setminus \bar{\Theta}_2) \cup \bar{\Theta}_1, G_\star)\.
\]

Then

\[
\bar{\theta}_{s'_1} < \bar{\theta}_{s_1} < \bar{\theta}_{s_2} < \bar{\theta}_{s'_2}.
\]

So Lemma 3 combined with the fact that \(B(G_\star, p)\) increases with \(p\) while \(B(G^*, p)\) decreases with \(p\) establishes

\[
B(G(s'_1), p(s'_1)) > B(G(s_i), p(s_i)), \quad i = 1, 2. \tag{22}
\]

In turn,

\[
m_{s_1} \bar{v}_{s_1} + m_{s_2} \bar{v}_{s_2} = m_{s_1} \cdot \frac{B(G(s_1), p(s_1)) + \kappa \bar{\theta}_{s_1}}{c} + m_{s_2} \cdot \frac{B(G(s_2), p(s_2)) + \kappa \bar{\theta}_{s_2}}{c}
\]

\[
= \frac{1}{c} \left[ m_{s_1} B(G(s_1), p(s_1)) + m_{s_2} B(G(s_2), p(s_2)) + \kappa \left(m_{s_1} \bar{\theta}_{s_1} + m_{s_2} \bar{\theta}_{s_2}\right) \right]
\]

\[
< \frac{1}{c} \left[ m_{s_1} B(G(s_1'), p(s_1')) + m_{s_2} B(G(s_2'), p(s_2')) + \kappa \left(m_{s_1} \bar{\theta}_{s_1'} + m_{s_2} \bar{\theta}_{s_2'}\right) \right]
\]

\[
= m_{s_1} \cdot \frac{B(G(s'_1), p(s'_1)) + \kappa \bar{\theta}_{s'_1}}{c} + m_{s_2} \cdot \frac{B(G(s'_2), p(s'_2)) + \kappa \bar{\theta}_{s'_2}}{c}
\]

\[
= m_{s_1} \bar{v}_{s'_1} + m_{s_2} \bar{v}_{s'_2}.
\]

\(^{19}\)In the former case, the grading rule is one of \(G^*\) and \(G_\star\).
The first and last equalities follow from (4). The second equality is a consequence of $m_{s_i'} = m_{s_i}$, for $i = 1, 2$, and $\sum_{i=1}^2 m_{s_i'} \theta_{s_i'} = \sum_{i=1}^2 m_{s_i} \theta_{s_i}$. The inequality is obtained from (22). We conclude that $S'$ generates greater effort than $S$, contradicting the fact that $M$ solves $(P_0')$. The proposition follows.

We now compare the solutions of $(P_0)$ and $(P_0')$. Let

$$s_1(z) := ([0, z], G^*) \land s_2(z) := ([z, 1], G^*)$$

and

$$S(z) := \begin{cases} 
{s_2(0)} & \text{if } z = 0 \\
{s_1(z), s_2(z)} & \text{if } 0 < z < 1 \\
{s_1(1)} & \text{if } z = 1
\end{cases}$$

Finally, define

$$R(z) := \begin{cases} 
\bar{c}_{s_2(0)} & \text{if } z = 0 \\
{m_{s_1(z)} \bar{c}_{s_1(z)} + m_{s_2(z)} \bar{c}_{s_2(z)}} & \text{if } 0 < z < 1 \\
\bar{c}_{s_1(1)} & \text{if } z = 1
\end{cases}$$

We can now state the main result of this appendix.

**Proposition D.3.** There exists $z$ such that a school system $S$ is part of a solution of $(P_0)$ if and only if $S = S(\hat{z})$, with either $\hat{z} = 0$, or $\hat{z} \in \arg\max_{z \in [0, 1]} R(z)$. A school system $S$ is part of a solution of $(P_0')$ if and only if $S = S(\hat{z})$, with $\hat{z} \in \arg\max_{z \in [0, 1]} R(z)$.

**Proof:** The second part of the proposition is an immediate consequence of Proposition D.2. We next prove the first part. The combination of Theorem 1 with Lemmata 4 and 5 shows that a school system $S$ is part of a solution of $(P_0)$ if and only if $S = S(\hat{z})$, with either $\hat{z} = 0$, $\hat{z} = 1$, or $\hat{z} \in (0, 1)$ and $c \bar{v}_{s_1(\hat{z})} - \kappa \theta_{s_1(\hat{z})} \leq c \bar{v}_{s_2(\hat{z})} - \kappa \theta_{s_2(\hat{z})}$. So all that remains to show is that $(\bar{v}_{s_2(z)} - \kappa \theta_{s_2(z)}) - (\bar{v}_{s_1(z)} - \kappa \theta_{s_1(z)})$ is a non-decreasing function of $z$.

Lemma 3 combined with the fact that $B(G_*, p)$ is increasing in $p$ shows that $p(s_2(z))$ increases with $z$. Hence, by (4), $\bar{c}_{s_2(z)} - \kappa \theta_{s_2(z)}$ increases with $z$ too. On the other hand, since $B(G_*, p)$ is decreasing in $p$, note that

$$B(G^*, \alpha \theta_{s_1(z)} + (1 - \alpha)x)$$
is a decreasing function of $z$ for every $x$. As $cx - \kappa \overline{\theta}_{s_1(z)}$ is also decreasing in $z$ for every $x$, we conclude from (4) that $\overline{\sigma}_{s_1(z)}^* - \kappa \overline{\theta}_{s_1(z)}$ decreases with $z$. Coupling the previous remarks proves the result we want.
References

Albano, G. L. and Lizzeri, A. (2001) Strategic Certification and Provision of Quality, *International Economic Review*, 42, 267–283.

Becker, W. E. and Rosen, S. (1992) The Learning Effect of Assessment and Evaluation in High School, *Economics of Education Review*, 11, 107–118.

Bergemann, D., Bonatti, A. and Smolin, A. (2018) The Design and Price of Information, *American Economic Review*, 108, 1–48.

Betts, J. (1998) The Impact of Educational Standards on the Level and Distribution of Earnings, *American Economic Review*, 88, 266–275.

Betts, J. (2011) The Economics of Tracking in Education, *Handbook of the Economics of Education*, 3, 341–381.

Bizzotto, J., Perez-Richet, E. and Vigier, A. (2021) Communication via Third Parties, *Available at SSRN 3530804*.

Bloedel, A. W. and Segal, I. (2021) Persuading a Rationally Inattentive Agent, Tech. rep., Working paper.

Blume, L. E., Brock, W. A., Durlauf, S. N. and Jayaraman, R. (2015) Linear Social Interactions Models, *Journal of Political Economy*, 123, 444–496.

Board, S. (2009) Monopolistic Group Design with Peer Effects, *Theoretical Economics*, 4, 89–125.

Board, S. and Meyer-ter Vehn, M. (2013) Reputation for Quality, *Econometrica*, 81, 2381–2462.

Boleslavsky, R. and Cotton, C. (2015) Grading Standards and Education Quality, *American Economic Journal: Microeconomics*, 7, 248–279.

Boleslavsky, R. and Kim, K. (2018) Bayesian Persuasion and Moral Hazard, *Available at SSRN 2913669*.

Brunello, G. and Checchi, D. (2007) Does School Tracking Affect Equality of Opportunity? New International Evidence, *Economic Policy*, 22, 781–861.
Chan, W., Hao, L. and Suen, W. (2007) A Signaling Theory of Grade Inflation, *International Economic Review*, **48**, 1065–1090.

Costrell, R. M. (1994) A Simple Model of Educational Standards, *American Economic Review*, pp. 956–971.

Doval, L. and Skreta, V. (2021) Constrained Information Design: Toolkit, *arXiv preprint arXiv:1811.03588*.

Dubey, P. and Geanakoplos, J. (2010) Grading Exams: 100, 99, 98 ... or A, B, C?, *Games and Economic Behavior*, **69**, 72–94.

Duflo, E., Dupas, P. and Kremer, M. (2011) Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya, *American Economic Review*, **101**, 1739–1774.

Esö, P. and Szentes, B. (2007) Optimal Information Disclosure in Auctions and the Handicap Auction, *The Review of Economic Studies*, **74**, 705–731.

Fang, H. and Moro, A. (2011) Theories of Statistical Discrimination and Affirmative Action: A Survey, *Handbook of social economics*, **1**, 133–200.

Gamoran, A. (2009) Tracking and Inequality: New Directions for Research and Practice, in *The Routledge International Handbook of the Sociology of Education*, Routledge.

Guo, Y. and Shmaya, E. (2019) The Interval Structure of Optimal Disclosure, *Econometrica*, **87**, 653–675.

Hanushek, E. and Wößmann, L. (2006) Does Educational Tracking Affect Performance and Inequality? Differences-in-Differences Evidence Across Countries, *Economic Journal*, **116**, C63–C76.

Hanushek, E. and Wößmann, L. (2011) The Economics of International Differences in Educational Achievement, *Handbook of the Economics of Education*, **3**, 89–200.

Jehiel, P. and Moldovanu, B. (2001) Efficient Design with Interdependent Valuations, *Econometrica*, **69**, 1237–1259.

Jehiel, P., Moldovanu, B. and Stacchetti, E. (1996) How (Not) to Sell Nuclear Weapons, *The American Economic Review*, pp. 814–829.
Jehiel, P., Moldovanu, B. and Stacchetti, E. (1999) Multidimensional Mechanism Design for Auctions with Externalities, *Journal of Economic Theory*, **85**, 258–293.

Kerr, S. P., Pekkarinen, T. and Uusitalo, R. (2013) School Tracking and Development of Cognitive Skills, *Journal of Labor Economics*, **31**, 577–602.

Kolotilin, A., Mylovanov, T., Zapechelnyuk, A. and Li, M. (2017) Persuasion of a Privately Informed Receiver, *Econometrica*, **85**, 1949–1964.

Li, H. and Shi, X. (2017) Discriminatory Information Disclosure, *American Economic Review*, **107**, 3363–85.

Lipnowski, E., Mathevet, L. and Wei, D. (2020) Attention Management, *American Economic Review: Insights*, **2**, 17–32.

Malamud, O. and Pop-Eleches, C. (2011) School Tracking and Access to Higher Education Among Disadvantaged Groups, *Journal of Public Economics*, **95**, 1538–1549.

Maurin, G. and McNally, S. (2012) The Effect of Tracking Students by Ability Into Different Schools a Natural Experiment, *Journal of Human Resources*, **47**, 684–721.

Ostrovsky, M. and Schwarz, M. (2010) Information Disclosure and Unraveling in Matching Markets, *American Economic Journal: Microeconomics*, **2**, 34–63.

Perez-Richet, E. and Skreta, V. (2022) Test Design under Falsification, *Econometrica*, **90**, 1109–1142.

Rayo, L. (2013) Monopolistic Signal Provision, *The BE Journal of Theoretical Economics*, **13**, 27–58.

Rosar, F. (2017) Test Design under Voluntary Participation, *Games and Economic Behavior*, **104**, 632–655.

Rothschild, C. and Scheuer, F. (2013) Redistributive Taxation in the Roy Model, *The Quarterly Journal of Economics*, **128**, 623–668.

Rothschild, C. and Scheuer, F. (2016) Optimal Taxation with Rent-seeking, *The Review of Economic Studies*, **83**, 1225–1262.
Saeedi, M. and Shourideh, A. (2020) Optimal Rating Design, *arXiv preprint arXiv:2008.09529*.

Tincani, M. M. (2018) Heterogeneous Peer Effects in the Classroom, *Manuscript, Dept. Econ., University College London*.

Wei, D. and Green, B. (2020) (Reverse) Price Discrimination with Information Design, *Available at SSRN 3263898*.

Wößmann, L. (2009) International Evidence on School Tracking: A Review, *CESifo DICE Report*, 1, 26–34.

Yamashita, T. and Sarkisian, R. (2021) Large Mechanism Design with Moment-based Allocation Externality, *TSE Working Papers 21-1241*.

Yamashita, T. and Zhu, S. (2021) Type-contingent Information Disclosure, *TSE Working Paper - 1242*.

Yang, K. H. (2022) Selling Consumer Data for Profit: Optimal Market-Segmentation Design and Its Consequences, *American Economic Review*, 112, 1364–93.

Zapechelnyuk, A. (2020) Optimal Quality Certification, *American Economic Review: Insights*, 2, 161–76.

Zubrickas, R. (2015) Optimal Grading, *International Economic Review*, 56, 751–776.
Online Appendix 1: Convex Costs

We assume in this appendix that effort costs are convex and type-independent. The effort cost of individual \( j \) is given by \( c(e_j) \), where \( c' > 0, c'' > 0, c''' \geq 0 \) and \( c'(1) > (1 - \alpha) b \).\(^2\) We denote with \( (P''_0) \) the associated designer problem.

Consider an arbitrary school \( s \) and type \( \theta \in \Theta_s \). The consistent effort level \( e^*_s(\theta) \) satisfies

\[
(1 - \alpha)\Delta(s)\left(q^+(s) - q^-(s)\right) = c'(e^*_s(\theta)). \tag{OA.1}
\]

As \( c' \) is invertible, (OA.1) uniquely defines \( e^*_s(\theta) \). This effort level does not depend on \( \theta \), hence

\[
e^*_s = e^*_s(\theta) = c'^{-1}\left((1 - \alpha)\Delta(s)\left(q^+(s) - q^-(s)\right)\right), \ \forall \theta \in \Theta_s. \tag{OA.2}
\]

Lastly, while Lemmata 2 and 3 hold verbatim we state below the equivalent of Lemma 1. The proofs of these three lemmata are omitted.

**Lemma OA.1.** Consider a school \( s \), and \( x \in [0, 1] \) satisfying

\[
B(G(s), \alpha \overline{\theta}_s + (1 - \alpha) x) > c'(x). \tag{OA.1}
\]

**Proof** that to any school \( s \) is associated a unique consistent profile of effort: The special case in which \( \Delta(s) = 0 \) is trivial. Suppose \( \Delta(s) > 0 \). We conclude from (OA.2) that, when potential employers believe that average students’ effort in school \( s \) is given by \( e^*_s \), exerting effort \( e^*_s \) maximizes the expected payoff of a type-\( \theta \) individual, for any \( \theta \in \Theta_s \).

The above observation shows that (OA.2) is a necessary and sufficient condition for an effort profile \( e^*_s(\cdot) \) to be consistent. The argument showing that this condition determines a unique profile of effort follows the steps of the proof for the baseline model (see Appendix A).

**Proposition OA.1.** If a school \( s \) maximizes the effort of its students, then either (i) \( G(s) = G^* \) or (ii) \( G(s) = G_* \). Furthermore, there exists a cutoff \( \theta^\dagger \) such that (i) holds whenever \( \overline{\theta}_s < \theta^\dagger \) and (ii) holds whenever \( \overline{\theta}_s > \theta^\dagger \).

**Proof:** The first part of the proposition is a consequence of Lemma 2. To prove the second part of the proposition, consider two schools, \( s_1 \) and \( s_2 \), sharing the same mean type, with grading rules \( G(s_1) = G^* \) and \( G(s_2) = G_* \). We claim that \( p(s_1) > 1/2 \) implies \( \overline{\theta}_{s_2} > \overline{\theta}_{s_1} \).

---

\(^2\)The last restriction will ensure that all students exert an interior level of effort.
Suppose \( p(s_1) > 1/2 \). Then

\[
B\left(G(s_2), \alpha \overline{\theta}_{s_2} + (1 - \alpha) \overline{e}_{s_1}^* \right) = B\left(G(s_2), p(s_1) \right) > B\left(G(s_1), p(s_1) \right) = c'\left(\overline{e}_{s_1}^* \right).
\]

The first equality follows from \( \theta_{s_1}^* = \theta_{s_2}^* \), the inequality is a consequence of Lemma 2 and \( p(s_1) > 1/2 \), and the last equality is obtained from (OA.2). We thus have \( B\left(G(s_2), \alpha \overline{\theta}_{s_2} + (1 - \alpha) \overline{e}_{s_1}^* \right) > c'\left(\overline{e}_{s_1}^* \right) \) and, by Lemma OA.1, \( \overline{e}_{s_2} > \overline{e}_{s_1}^* \).

Now, Lemma 3 implies the existence of a cutoff \( \theta^\dagger \) such that \( p(s) > 1/2 \) whenever \( \overline{\theta}_s > \theta^\dagger \) and \( G(s) = G^\star \). Combining the first part of the proposition with the previous claim, this observation shows that if a school \( s \) with \( \theta_s > \theta^\dagger \) maximizes the effort of its students, then \( G(s) = G_\star \). One shows analogously that if a school \( s \) with \( \theta_s < \theta^\dagger \) maximizes the effort of its students, then \( G(s) = G^\star \).

In the rest of this appendix, \( \theta^\dagger \) will refer to the cutoff defined in Proposition OA.1.

**Proposition OA.2.** If two schools share the same informative grading rule, then merging them increases total effort.

**Proof:** Consider the school system \( S = \{s_1, s_2\} \) such that \( s_1 \) and \( s_2 \) share the same informative grading rule, \( G \) say. Let \( S' = \{s_3\} \), where \( s_3 = ([0,1], G) \). We have to show that

\[
\overline{e}_{s_3}^* > m_{s_1} \overline{e}_{s_1}^* + m_{s_2} \overline{e}_{s_2}^*.
\]

As \( m_{s_2} = 1 - m_{s_1} \), \( \overline{\theta}_{s_3} = m_{s_1} \overline{\theta}_{s_1} + m_{s_2} \overline{\theta}_{s_2} \), \( B(G, \cdot) \) is strictly concave, and \( c \) is strictly convex, using (OA.2) yields

\[
B\left(G, \alpha \overline{\theta}_{s_3} + (1 - \alpha)(m_{s_1} \overline{e}_{s_1}^* + m_{s_2} \overline{e}_{s_2}^*) \right) > m_{s_1} B(G, p(s_1)) + m_{s_2} B(G, p(s_2))
\]

\[
= m_{s_1} c'(\overline{e}_{s_1}^*) + m_{s_2} c'(\overline{e}_{s_2}^*)
\]

\[
> c'(m_{s_1} \overline{e}_{s_1}^* + m_{s_2} \overline{e}_{s_2}^*).
\]

Using Lemma 1 now establishes (OA.3). \[\blacksquare\]

**Theorem OA.1.** Any solution of \((P''_0)\) comprises a school system which either contains a single school, or exhibits a two-tier structure. In the latter case, the bottom-tier school has grading rule \( G^\star \), while the top-tier school has grading rule \( G_\star \). \cite{21}

\[\text{In the former case, the grading rule is one of } G^\star \text{ and } G_\star.\]
We prove the theorem by way of two propositions. Let \((P''_3)\) denote the auxiliary problem of the designer obtained from \((P'_{0})\) by adding to it the constraint that every \(\Theta_s\) has to be an interval.

**Proposition OA.3.** Any solution of \((P''_3)\) comprises a school system which either contains a single school, or exhibits a two-tier structure. In the latter case, the bottom-tier school has grading rule \(G^\ast\), while the top-tier school has grading rule \(G_\ast\).

**Definition OA.1.** A school system \(S\) is regular* if for any pair of schools \(s_1\) and \(s_2\) such that \(\underline{e}_s < \overline{e}_s\), we have \(\sup \Theta_s \leq \inf \Theta_s\).

**Remark OA.1.** Any regular* school system \(S\) is characterized by unique cutoffs \(\hat{\theta}_0 = 0 < \hat{\theta}_1 < \cdots < \hat{\theta}_{I-1} < 1 = \hat{\theta}_I\) such that to every \(s \in S\) is associated a unique index, say \(i(s)\), satisfying the following properties

(i) \(\Theta_s \subseteq [\hat{\theta}_{i(s)-1}, \hat{\theta}_{i(s)}]\),

(ii) \(i(s') > i(s) \Leftrightarrow \overline{e}_{s'} > \overline{e}_s\).

**Lemma OA.2.** Any feasible mechanism comprises a regular* school system.

**Proof:** Pick a feasible \(M = (S, t)\). We have

\[
U(\theta, e_\ast(\theta), s) = q^-(s) + \left[ g_0(s) + (\alpha \theta + (1 - \alpha) e_\ast(\theta)) \Delta(s) \right] (q^+(s) - q^-(s)) - c(e_\ast(\theta)) - t(s) \\
= q^-(s) + \left[ g_0(s) \Delta(s) + \alpha \theta \right] \frac{c'(\overline{e}_s)}{1 - \alpha} + \overline{e}_s c'(\overline{e}_s) - c(\overline{e}_s) - t(s) \\
= \bar{\xi}(s) + \frac{\alpha}{1 - \alpha} \theta c'(\overline{e}_s),
\]

where

\[
\bar{\xi}(s) := q^-(s) + \frac{g_0(s) c'(\overline{e}_s)}{\Delta(s)} + \overline{e}_s c'(\overline{e}_s) - c(\overline{e}_s) - t(s).
\]

As \(c'\) is increasing, standard mechanism design arguments complete the proof.

**Lemma OA.3.** Any regular* school system forms part of a feasible mechanism.

**Lemma OA.4.** If \(G(s) = G_\ast\), then \(\overline{e}_s\) is an increasing function of \(\overline{\theta}_s\).

The proofs of Lemma OA.3 and Lemma OA.4 are almost identical, respectively, to the proofs of Lemma 5 and Lemma 6.
**Proof of Proposition OA.3:** Suppose some M = (S, t) with |S| ≥ 2 solves \((P'')_3\). Let \(S = \{s_1, \ldots, s_I\}\), with schools labeled so as to satisfy sup \(\Theta_{s_i} \leq \inf \Theta_{s_k}\) for every \(i < k\). Then Lemma OA.2 gives
\[
\overline{e}_{s_1} \leq \cdots \leq \overline{e}_{s_I}. \tag{OA.4}
\]
Furthermore, by the same arguments used in the proof of Proposition OA.3, we have \(G(s_I) = G_*\) and \(\overline{\theta}_{s_I} > \theta^* > \overline{\theta}_{s_{I-1}}\). Next, reason by contradiction and suppose either \(I > 2\), or \(I = 2\) but \(G(s_1) \neq G_*\). Consider the school system \(S' = \{s'_1, s'_2\}\) where
\[
s'_1 = \left( \bigcup_{i=1}^{I-1} \Theta_{s_i}, G_* \right), \quad s'_2 = s_I.
\]
By Propositions OA.1 and OA.2,
\[
\overline{e}_{s'_1} > \frac{1}{I-1} \sum_{i=1}^{I-1} m_{s_i} \overline{e}_{s_i}.
\]
So \(S'\) cannot be regular, for otherwise applying Lemma OA.3 would yield a contradiction with the optimality of \(M\) for problem \((P''_3)\). Hence:
\[
\overline{e}_{s'_1} > \overline{e}_{s'_2}. \tag{OA.5}
\]
Now, given \(0 \leq z < 1 - \inf \Theta_{s'_2}\), consider the school system \(S''(z) = \{s''_1(z), s''_2(z)\}\), where
\[
s''_1(z) = \left( [0, \inf \Theta_{s'_2} + z], G_* \right), \quad s''_2(z) = \left( \inf \Theta_{s'_2} + z, 1 \right], G_* \right).
\]
The difference \(D(z) = \overline{e}_{s''_1(z)} - \overline{e}_{s''_2(z)}\) is continuous in \(z\), and, by (OA.5), satisfies \(D(0) > 0\). So by virtue of the intermediate value theorem, either \(D(z) > 0\) for all \(z \in (0, 1 - \inf \Theta_{s'_2})\), or there exists a \(\tilde{z} \in (0, 1 - \inf \Theta_{s'_2})\) such that \(D(\tilde{z}) = 0\). Suppose that the second case holds (the arguments are similar for the first case, and therefore omitted). Then
\[
m_{s''_1(\tilde{z})} \overline{e}_{s''_1(\tilde{z})} + m_{s''_2(\tilde{z})} \overline{e}_{s''_2(\tilde{z})} = \overline{e}_{s''_1(\tilde{z})} > \overline{e}_{s''_2(\tilde{z})} = \overline{e}_{s'_2}, \geq \sum_{i=1}^{I} m_{s_i} \overline{e}_{s_i},
\]
where the first equality follows from \(D(\tilde{z}) = 0\), the strict inequality from Lemma OA.4, and the weak inequality from (OA.4). We thus obtain \(m_{s''_1(\tilde{z})} \overline{e}_{s''_1(\tilde{z})} + m_{s''_2(\tilde{z})} \overline{e}_{s''_2(\tilde{z})} > \sum_{i=1}^{I} m_{s_i} \overline{e}_{s_i}\). But notice that, by virtue of Lemma OA.3, \(S''(\tilde{z})\) forms part of a feasible mechanism. The last
inequality therefore contradicts our hypothesis that $M$ solves $(P'')$. ■

**Lemma OA.5.** Consider a school $s$ with $G(s) = G^*$. Then $\overline{e}_s$ is a decreasing function of $\overline{\theta}_s$.

**Proof:** Note to begin with that $B(G^*, \cdot)$ is a decreasing function. Thus, if $G(s) = G^*$, Lemma 3 tells us that $B(G(s), p(s))$ is decreasing in $\overline{\theta}_s$. Lemma 1 completes the proof. ■

The following proposition concludes the proof of Theorem OA.1.

**Proposition OA.4.** Any school system $S$ forming part of a solution of $(P'')$ is such that $\Theta_s$ is almost an interval, for all $s \in S$.

**Proof:** Let $S$ be part of a mechanism solving $(P'')$. By Lemma OA.2, $S$ is regular*. Let $\{\hat{\theta}_i\}_{i=0}^I$ be the set of cutoffs characterizing $S$ (see Remark OA.1). Suppose $I = 1$ (the arguments are similar for $I > 1$), that is,

$$\overline{\theta}_s = \overline{\theta}_{s_k}, \quad \forall s, s_k \in S. \quad (OA.6)$$

Below, let $S_- := \{s \in S : \overline{\theta}_s < \theta_s^\uparrow\}, S_+ := \{s \in S : \overline{\theta}_s > \theta_s^\uparrow\},$ and $S_{\geq} := S \setminus S_-.$

Now reason by contradiction, and suppose that not every school $s \in S$ is such that $\Theta_s$ is almost an interval. We can then find two schools in $S$, say $s_1$ and $s_2$, with $\overline{\theta}_{s_1} \leq \overline{\theta}_{s_2}$, as well as subsets $\tilde{\Theta}_i \subset \Theta_s$, for $i = 1, 2$ satisfying: (a) $\int_{\tilde{\Theta}_1} dF(\theta) = \int_{\tilde{\Theta}_2} dF(\theta) > 0$, and (b) $\sup \tilde{\Theta}_2 \leq \inf \tilde{\Theta}_1$. We treat below the case in which $s_1 \in S_- \subset S$ and $s_2 \in S_+.$

Define $S' = \{s'_1, s'_2\}$ as follows:

$$s'_1 = \left(\cup_{s \in S_-} \Theta_s, G^*\right), \quad s'_2 = \left(\cup_{s \in S_+} \Theta_s, G^*\right).$$

Using Propositions OA.1 and OA.2 shows

$$\overline{e}'_s \geq \frac{1}{m_{s_1}} \sum_{s \in S_-} m_s \overline{e}_s, \quad \overline{e}'_s \geq \frac{1}{m_{s_2}} \sum_{s \in S_+} m_s \overline{e}_s, \quad (OA.7)$$

Next, define $S'' = \{s''_1, s''_2\}$ by

$$s''_1 = \left(\left(\Theta_{s_1} \setminus \tilde{\Theta}_1\right) \cup \tilde{\Theta}_2, G^*\right), \quad s''_2 = \left(\left(\Theta_{s_2} \setminus \tilde{\Theta}_2\right) \cup \tilde{\Theta}_1, G^*\right).$$

Note that, by construction,

$$\overline{\theta}_{s''_1} < \overline{\theta}_{s'_1} < \overline{\theta}_{s'_2} < \overline{\theta}_{s''_2}. \quad (OA.8)$$
So Lemma 6 and Lemma OA.5 give, respectively, $\bar{v}_{s_2} > \bar{v}_{s_1}$ and $\bar{v}_{s_2} > \bar{v}_{s_1}$. In turn, using (OA.7) yields

$$\bar{v}_{s_1} > \frac{1}{m_{s_1}} \sum_{s \in S_1} m_s \bar{v}_s, \quad \bar{v}_{s_2} > \frac{1}{m_{s_2}} \sum_{s \in S_2} m_s \bar{v}_s. \quad (OA.9)$$

By virtue of Lemma 5, if $c\bar{v}_{s_2} - \kappa \bar{\theta}_{s_2} = c\bar{v}_{s_1} - \kappa \bar{\theta}_{s_1}$ then $S''$ would form part of a feasible mechanism. But this cannot be, as (OA.9) would then contradict the optimality of $S$ for problem (P_0'). So $c\bar{v}_{s_2} - \kappa \bar{\theta}_{s_2} \neq c\bar{v}_{s_1} - \kappa \bar{\theta}_{s_1}$. Say $c\bar{v}_{s_1} - \kappa \bar{\theta}_{s_1} > c\bar{v}_{s_2} - \kappa \bar{\theta}_{s_2}$ (otherwise, just switch the roles of the two schools in what follows), and let $S''' = \{s_1', s_2''\}$, with

$$s_1' = \left(\Theta_{s_1}, (1 - b, \Delta''')\right), \quad s_2'' = s_2'',$

and $\Delta'''$ chosen so as to guarantee $\bar{v}_{s_1'} = \bar{v}_{s_2''}$. Then

$$m_{s_1'} \bar{v}_{s_1'} + m_{s_2''} \bar{v}_{s_2''} = \bar{v}_{s_2''} = \bar{v}_{s_2'} > \frac{1}{m_{s_2}} \sum_{s \in S_2} m_s \bar{v}_s = \sum_{s \in S} m_s \bar{v}_s.$$

The strict inequality follows from (OA.9), and the last equality is a consequence of (OA.6). Since, by virtue of Lemma OA.3, $S'''$ forms part of a feasible mechanism, the inequality $m_{s_1'} \bar{v}_{s_1'} + m_{s_2''} \bar{v}_{s_2''} > \sum_{s \in S} m_s \bar{v}_s$ contradicts the fact that $S$ is part of a mechanism solving (P_0').

$\blacksquare$
Online Appendix 2: Social Welfare

In this appendix, we study the problem of a designer choosing a feasible mechanism with a view to maximizing social welfare. Let \((P_0'')\) denote the auxiliary problem of the designer obtained from \((P_0)\) by assuming that the designer aims to maximize social welfare:

\[
\max_{M} W(M) := \sum_{s \in S} \int_{\Theta_s} ((1 - \alpha)e_s^*(\theta) - c(e_s^*(\theta), \theta))dF(\theta), \quad \text{s.t. } M \text{ is feasible.}
\]

(\(P_0''\))

Using (5), we reformulate the planner’s objective:

\[
W(M) = \sum_{s \in S} \left[ (1 - \alpha)m_sk - \int_{\Theta_s} \left( \frac{c}{2}(e_s^* + \frac{\kappa}{c}(\theta - \overline{\theta}_s)) \right)^2 - \kappa\theta(e_s^* + \frac{\kappa}{c}(\theta - \overline{\theta}_s)) \right]dF(\theta),
\]

and, after straightforward calculations,

\[
W(M) = \frac{1}{c} \sum_{s \in S} m_s \left[ (1 - \alpha)(c\overline{e}_s - \kappa\overline{\theta}_s) - \frac{1}{2}(c\overline{e}_s - \kappa\overline{\theta}_s)^2 \right] + r',
\]

where \(r'\) is a constant independent of \(M\). The problem of a planner maximizing social welfare is thus the same as the problem of a planner maximizing

\[
\tilde{W}(M) := \sum_{s \in S} m_s v(c\overline{e}_s - \kappa\overline{\theta}_s),
\]

where \(v(x) := (1 - \alpha)x - \frac{1}{2}x^2\).

We next state two simple yet key lemmata. The first shows that \(v\) is increasing over the relevant range, the second shows that whenever merging two schools increases overall effort, it also increases welfare.

**Lemma OA.6.** Consider two schools, denoted \(s_1\) and \(s_2\), such that \(c\overline{e}_{s_1} - \kappa\overline{\theta}_{s_1} > c\overline{e}_{s_2} - \kappa\overline{\theta}_{s_2}\); then \(v(c\overline{e}_{s_1} - \kappa\overline{\theta}_{s_1}) > v(c\overline{e}_{s_2} - \kappa\overline{\theta}_{s_2})\).

**Proof:** Function \(v\) is increasing as long as \(x < 1 - \alpha\). Yet, using (4) shows \(c\overline{e}_s - \kappa\overline{\theta}_s = B(G(s), p(s)) < (1 - \alpha)b < 1 - \alpha\) for any school \(s\).

---

\(^{22}\)Social welfare is given by \(\sum_{s \in S} \left[ a\overline{\theta}_s + \int_{\Theta_s} ((1 - \alpha)e_s^*(\theta) - c(e_s^*(\theta), \theta))dF(\theta) \right]\). Yet the term \(\sum_{s \in S} a\overline{\theta}_s\) is a constant and can be disregarded.
Lemma OA.7. Consider three schools, denoted $s_1$, $s_2$ and $s_3$, such that (i) $\Theta_{s_1} \cap \Theta_{s_2} = \emptyset$, (ii) $\Theta_{s_3} = \Theta_{s_1} \cup \Theta_{s_2}$, and (iii) $m_{s_3} \bar{v}_{s_3} \geq m_{s_1} \bar{v}_{s_1} + m_{s_2} \bar{v}_{s_2}$. Then

$$m_{s_3} v(\bar{c}_{s_3} - \kappa \bar{\theta}_{s_3}) > m_{s_1} v(\bar{c}_{s_1} - \kappa \bar{\theta}_{s_1}) + m_{s_2} v(\bar{c}_{s_2} - \kappa \bar{\theta}_{s_2}).$$

Proof: Let schools $s_1$, $s_2$ and $s_3$ satisfy the properties listed in the lemma. Then

$$m_{s_3} \bar{v}_{s_3} \geq m_{s_1} \bar{v}_{s_1} + m_{s_2} \bar{v}_{s_2} \iff m_{s_3} (\bar{c}_{s_3} - \kappa \bar{\theta}_{s_3}) \geq m_{s_1} (\bar{c}_{s_1} - \kappa \bar{\theta}_{s_1}) + m_{s_2} (\bar{c}_{s_2} - \kappa \bar{\theta}_{s_2}) \Rightarrow$$

$$m_{s_3} v(\bar{c}_{s_3} - \kappa \bar{\theta}_{s_3}) > m_{s_1} v(\bar{c}_{s_1} - \kappa \bar{\theta}_{s_1}) + m_{s_2} v(\bar{c}_{s_2} - \kappa \bar{\theta}_{s_2}),$$

where the first arrow follows from $m_{s_3} \bar{\theta}_{s_3} = m_{s_1} \bar{\theta}_{s_1} + m_{s_2} \bar{\theta}_{s_2}$, and the second arrow follows from strict concavity of $v(\cdot)$.

The following proposition is an immediate corollary of Lemma OA.6.

Proposition OA.5. If a school $s$ maximizes the welfare of its students, then either (i) $G(s) = G^*$ or (ii) $G(s) = G_\ast$. Furthermore, there exists a cutoff $\theta^\circ$ such that (i) holds whenever $\bar{\theta}_s < \theta^\circ$ and (ii) holds whenever $\bar{\theta}_s > \theta^\circ$.

In what follows $\theta^\circ$ will refer to the cutoff defined in Proposition OA.5. The following proposition is an immediate corollary of Lemma OA.7.

Proposition OA.6. If two schools share the same informative grading rule, then merging them increases welfare.

We are now ready to state the main result of this appendix.

Theorem OA.2. Any solution of $\textbf{(P''')}$ comprises a school system which either contains a single school, or exhibits a two-tier structure. In the latter case, the bottom-tier school has grading rule $G^*$, while the top-tier school has grading rule $G_\ast$.\[23\]

Let $\textbf{(P''')}$ denote the auxiliary problem of the designer obtained from $\textbf{(P''')}$ by adding to it the constraint that every $\Theta_s$ has to be an interval. The following proposition summarizes the first step of our proof of Theorem OA.2.

\[23\text{In the former case, the grading rule is one of } G^* \text{ and } G_\ast.\]
Proposition OA.7. Any solution of \((P_3^{''''})\) comprises a school system which either contains a single school, or exhibits a two-tier structure. In the latter case, the bottom-tier school has grading rule \(G^*\), while the top-tier school has grading rule \(G_\ast\).

Proof of Proposition OA.7: Suppose some mechanism \(M = (S, t)\) with \(|S| \geq 2\) solves \((P_3^{''''})\). Let \(S = \{s_1, \ldots, s_I\}\), with schools labeled so as to satisfy \(\sup \Theta_{s_i} \leq \inf \Theta_{s_k}\) for every \(i < k\). Then Lemma 4 gives

\[ c\bar{e}_{s_1} - \kappa \bar{\theta}_{s_1} \leq \cdots \leq c\bar{e}_{s_I} - \kappa \bar{\theta}_{s_I}. \]  

(OA.10)

Arguments akin to those used in the proof of Proposition 3 ensure that \(\bar{\theta}_{s_I} > \theta^\circ > \bar{\theta}_{s_{I-1}}\) and \(G(s_I) = G_\ast\). Next, reason by contradiction and suppose either \(I > 2\), or \(I = 2\) but \(G(s_1) \neq G^*\). Consider the school system \(S' = \{s'_1, s'_2\}\) where

\[ s'_1 = \left( \bigcup_{i=1}^{I-1} \Theta_{s_i}, G^* \right), \quad s'_2 = s_I. \]

By Propositions OA.5 and OA.6, and Lemma OA.7

\[ \sum_{i=1}^{I-1} m_{s_i} v \left( c\bar{e}_{s'_1} - \kappa \bar{\theta}_{s'_1} \right) > \sum_{i=1}^{I-1} m_{s_i} v \left( c\bar{e}_{s_i} - \kappa \bar{\theta}_{s_i} \right). \]

So \(S'\) cannot be regular, for otherwise applying Lemma 5 would yield a contradiction with the optimality of \(M\) for problem \((P_3^{''''})\). Hence: \(c\bar{e}_{s'_1} - \kappa \bar{\theta}_{s'_1} > c\bar{e}_{s'_2} - \kappa \bar{\theta}_{s'_2}\). Now, given \(0 \leq z < 1 - \inf \Theta_{s'_2}\), consider the school system \(S''(z) = \{s''_1(z), s''_2(z)\}\), where

\[ s''_1(z) = \left( [0, \inf \Theta_{s'_2} + z], G^* \right), \quad s''_2(z) = \left( \inf \Theta_{s'_2} + z, 1 \right], G_\ast \right). \]

The difference

\[ D(z) = \left( c\bar{e}_{s''_1(z)} - \kappa \bar{\theta}_{s''_1(z)} \right) - \left( c\bar{e}_{s''_2(z)} - \kappa \bar{\theta}_{s''_2(z)} \right) \]

is continuous in \(z\) and satisfies \(D(0) > 0\). So by virtue of the intermediate value theorem, either \(D(z) > 0\) for all \(z \in (0, 1 - \inf \Theta_{s'_2})\), or there exists \(\tilde{z} \in (0, 1 - \inf \Theta_{s'_2})\) such that \(D(\tilde{z}) = 0\).
Suppose that the second case holds. Then

\[
m_{s_1''}(z)v(\hat{c}\hat{e}'_{s_1''}(z) - \kappa \hat{\theta}_{s_1''}(z)) + m_{s_2''}(z)v(\hat{c}\hat{e}'_{s_2''}(z) - \kappa \hat{\theta}_{s_2''}(z)) = v(\hat{c}\hat{e}'_{s_2''}(z) - \kappa \hat{\theta}_{s_2''}(z)) > v(\hat{c}\hat{e}'_{s_2'} - \kappa \hat{\theta}_{s_2'}) \\
\geq \sum_{i=1}^{I} m_{s_i}v(\hat{c}\hat{e}_{s_i} - \kappa \hat{\theta}_{s_i}),
\]

where the first equality follows from \(D(\tilde{z}) = 0\), the strict inequality from Lemma 6, and the weak inequality from (OA.10). But notice that, by virtue of Lemma 5, \(S''(\tilde{z})\) forms part of a feasible mechanism. The last inequality contradicts our hypothesis that \(M\) solves \((P_{0''}'')\).

The following proposition summarizes the last step of our proof of Theorem OA.2.

**Proposition OA.8.** Any school system \(S\) forming part of a solution of \((P_{0''}'')\) is such that \(\Theta_s\) is almost an interval, for all \(s \in S\).

**Proof of Proposition OA.8:** Let \(S\) be part of a mechanism solving \((P_{0''}'')\). By Lemma 4, \(S\) is a regular school system. Let \(\{\hat{\theta}_i\}_{i=0}^I\) be the set of cutoffs characterizing \(S\) (see Remark 1). Suppose \(I = 1\) (the arguments are similar for \(I > 1\), and therefore omitted), that is,

\[
\hat{c}\hat{e}'_{s_i} - \kappa \hat{\theta}_{s_i} = \hat{c}\hat{e}'_{s_k} - \kappa \hat{\theta}_{s_k}, \quad \forall s_i, s_k \in S.
\]

Below, let \(S_< := \{s \in S : \hat{\theta}_s < \theta^o\}\) and \(S_> := \{s \in S : \hat{\theta}_s > \theta^o\}\).

Now reason by contradiction, and suppose that not every schools \(s \in S\) is such that \(\Theta_s\) is almost an interval. We can then find two schools in \(S\), say \(s_1\) and \(s_2\), with \(\hat{\theta}_{s_1} \leq \hat{\theta}_{s_2}\), as well as subsets \(\hat{\Theta}_i \subset \Theta_{s_i}\) for \(i = 1, 2\) satisfying: (a) \(\int_{\hat{\theta}_1} dF(\theta) = \int_{\hat{\theta}_2} dF(\theta) > 0\), and (b) \(\sup \hat{\Theta}_2 \leq \inf \hat{\Theta}_1\). Lemma OA.7 implies that the case in which \(s_1 \in S_<\) and \(s_2 \in S_>\) can be dealt with using the arguments given in the proof of Proposition 4. The other cases are analogous and therefore omitted. ■