Pay for Percentile: Web Appendix

Gadi Barlevy

Federal Reserve Bank of Chicago

Derek Neal

University of Chicago and NBER

July 25, 2011

Appendix A

We argue that our pay for percentile scheme is advantageous because it is robust against certain forms of coaching and scale manipulation. Because we do not model coaching and scale manipulation explicitly in our main analyses, this Appendix provides a more precise explanation of what we mean by coaching and scale manipulation. To simplify notation, we assume $g(a) = 0$ for all a, so that

$$a'_{ij} = t_j + \alpha e_{ij} + \varepsilon_{ij} \tag{1}$$

We can think of this as a special case of our environment in which all students enjoy the same baseline achievement level.

Coaching

First, we model coaching using a simple version of Holmstrom and Milgrom's (1991) multi-tasking model. Here, we define coaching as the time that teachers devote to memorization of the answers to questions on previous exams, instruction on test taking techniques that are specific to the format of previous exams, or test taking practice sessions that involve repeated exposure to previous exam questions. These activities may create human capital, but we assume that these activities are socially wasteful because they create less human capital per hour than classroom time devoted to effective comprehensive instruction. However, teachers still have an incentive to coach if coaching has a large enough effect on test scores.

Suppose that in addition to the effort choices $\mathbf{e}_j$ and t_j, teachers can also engage in coaching. Denote the amount of time teacher j spends on this activity by τ_j. To facilitate exposition, assume τ_j and t_j are perfectly substitutable in the teacher's cost of effort, since both represent time working with the class as a whole. That is, the cost of effort for teacher j is given by $C(\mathbf{e}_j, t_j + \tau_j)$, where $C(\cdot,\cdot)$ is once again assumed to be convex.

We allow τ_j to affect human capital, but with a coefficient θ that is less than 1, i.e., we replace (**??**) with

$$a'_{ij} = t_j + \theta\tau_j + \alpha e_{ij} + \varepsilon_{ij}$$

Hence, τ_j is less effective in producing human capital than general classroom instruction t_j. At the same time, suppose τ_j helps to raise the test score for a student of a given achievement level by improving their chances of getting specific items correct on a test. That is,

$$s'_{ij} = m(a'_{ij} + \mu\tau_j)$$

where $\mu \geq 0$ reflects the role of coaching in improving test scores that has no impact on the student's actual human capital.

For any compensation scheme that is increasing in s'_{ij}, teachers will choose to coach rather than teach if $\theta + \mu > 1$. It seems reasonable to assert that μ is an increasing function of the fraction of items on a given assessment that are repeated items from previous assessments. The return to coaching kids on the answers to specific questions or how to deal with questions in specific formats should be increasing in the frequency with which these specific questions and formats appear on the future assessments. Because coaching is socially wasteful, the education authority would like to make each year's test so different from previous exams that $\theta + \mu < 1$. However, any attempt to minimize the use of previous items will make it more difficult to place scores on a common scale across assessments, and in the limit, standard psychometric techniques cannot be used to equate two assessments given at different points in time, and thus presumably to populations with different distributions of human capital, if these assessments contain no common items. Thus, in practice, scale-dependent compensation schemes create opportunities for coaching because they require the repetition of items over time. Scale-invariant compensation schemes, such as our pay for percentile system, can be used to reduce coaching because these schemes can employ a series of assessments that contain no repeated items.

Although we chose to interpret τ_j as coaching, teachers may try to raise test scores by engaging in other socially inefficient activities. The ordinal system we propose can help education authorities remove incentives for teachers to coach students concerning specific questions that are likely to appear on the next assessment. However, in any assessment-based performance pay system, individual teachers can often increase their reward pay by taking actions that are hidden from the education authority. The existing empirical literature suggests that these actions take many forms, e.g. manipulating the population of students tested and changing students' answers before their tests are graded.

Our scheme is less vulnerable to coaching, but no less vulnerable to these types of distortions.

Scale Manipulation

Assume that the education authority contracts with a testing agency to test students and report the results of the tests. The testing agency knows the true mapping between test results and human capital, and thus the authority can contract with the testing agency to report scores in units of human capital, so that

$$s = m(a) = a$$

The educational authority then announces the rule by which teachers will be compensated. This rule is a mapping from the set of all reported test scores to the set of incomes for each teacher. Let $x_j(\mathbf{s}'_1, ..., \mathbf{s}'_J)$ equal teacher j's income under the announced rule, where $\mathbf{s}'_j = (s_{1j}, ..., s_{Nj})$ is the vector of scores of all students in class j at the end of the year. We argue in Section II that, if the educational authority knows that the scores are scaled in units of human capital, it can maximize total surplus from the production of human capital by adopting the following relative performance pay scheme

$$x_j(\mathbf{s}'_1, ..., \mathbf{s}'_J) = NX_0 + R\sum_{i=1}^{N}\left[s'_{ij} - \frac{1}{J}\sum_{j=1}^{J} s'_{ij}\right] \tag{2}$$

where $s' = m(a') = a'$.

We assume that, after $x(\cdot)$ is announced, the teachers have opportunity to collectively lobby the testing agency to report scores on a different scale, which we restrict to take the form

$$s = \widehat{m}(a) = \hat{\lambda}a + \hat{\phi}$$

where $\hat{\lambda} > 0$, which implies that any manipulation must preserve ordering. Our concern is scaling, not errors in measurement. The teachers or their union may engage in this type of manipulation by lobbying the testing agency to choose sets of questions that are less discriminating and do not capture the true extent of human capital differences among students, or they may lobby the testing agency to falsely equate scores from current assessments to some established baseline scale.

Note that (**??**) is robust against efforts to inflate scores by choosing $\hat{\phi} > 0$ because payments are based on relative performance. However, the teachers can still benefit from convincing the testing agency to choose $\hat{\lambda} < 1$. Given this scheme or other symmetric, relative compensation schemes, all teachers face the same incentives and will thus put in the same effort. This implies that, if the testing agency reports $\hat{m}(a)$ instead of $m(a)$, the expected teacher payoff from (**??**) equals $NX_0 - C(\hat{\mathbf{e}}_j, \hat{t}_j)$, where $(\hat{\mathbf{e}}_j, \hat{t}_j)$ denotes the common level of effort teachers choose in response to the scale that they have convinced the testing agency to employ. Any manipulation of the scale that induces teachers to coordinate on a lower common level of effort will make them all better off. To see that teachers can coordinate on less effort by manipulating $\hat{\lambda}$, note that under (**??**), each teacher would take the effort decisions of all other teachers as given and would face the following problem when choosing their own effort,

$$\max_{\mathbf{e}_j, t_j} NX_0 + \sum_{i=1}^{N} R\left[\hat{\lambda}(t_j + \alpha e_{ij}) - \text{constant}\right] - C\left(\mathbf{e}_j, t_j\right)$$

and thus $(\hat{\mathbf{e}}_j, \hat{t}_j)$ are decreasing functions of $\hat{\lambda}$.

Of course, just because teachers have an incentive to manipulate $\hat{\lambda}$, does not mean that they can do so without being detected by the authority. Recall that, in equilibrium, all teachers put in the same effort, and thus, any variation in scores is due to ε_{ij}. Thus, if the education authority knows the distribution of ε_{ij}, it can detect whether $\hat{\lambda} = 1$ by comparing the standard deviation of test scores with the standard deviation of ε_{ij}.

However, if the education authority is uncertain about the distribution of ε_{ij} but teachers know the distribution, the education authority may be unable to detect $\hat{\lambda} < 1$. Suppose there are two states of the world, each equally likely, which differ in the dispersion of the shock term, ε_{ij}. In the first state, the shock is ε_{ij}, as above, and in the second state the shock is scaled up by a constant $\sigma > 1$. If teachers observe that they are in the second state of the world, they can set $\hat{\lambda} = 1/\sigma$ (and choose $\hat{\phi}$ appropriately) so that the distribution of test scores is identical to the distribution of test scores in the first state of the world when $\widehat{\lambda} = 1$. Thus, under the relative pay for performance scheme above, teachers could manipulate the scale without being detected, and they would benefit from this manipulation.

By contrast, our pay-for-percentile scheme has the advantage of not being vulnerable to this type of manipulation. By construction, changing the scale has no effect on the compensation teachers would earn under pay for percentile. Our scheme does require identifying which of the two states of the world we are in, since $h\left(0\right)$ will differ in these states. However, the procedure we suggest for recovering $h\left(0\right)$ works in either state.[1]

Appendix B: Proofs

Proposition 1: *Let $\widetilde{\varepsilon}_{ij}$ denote a random variable with mean zero, and let $\varepsilon_{ij} = \sigma\widetilde{\varepsilon}_{ij}$. There exists $\overline{\sigma}$ such that $\forall\sigma > \overline{\sigma}$, in a two-teacher contest, both*

[1] Note that there may be scale-dependent schemes that still elicit first best. For example, if the education authority offers a higher base pay whenever the dispersion of scores is high, it can provide incentives that dissuade teachers from distorting the scale. Since a change in base pay does not affect the incentive to put in effort, such a scheme can elicit first best effort.

teachers choosing the socially optimal effort levels $(\mathbf{e}^*, t^*)$ *is a pure strategy Nash equilibrium.*

Proof of Proposition 1: Our proof will be for the general production function $a'_{ij} = g_i(\mathbf{a},, t_j, \mathbf{e}_j)$, where $g_i(\cdot)$ is concave in $(t_j, \mathbf{e}_j)$. Note that this includes the separable production function in Section I as a special case. While the arguments presented here address two teacher contests, at the end of the proof, we discuss how they generalize to the K teacher case.

Define $\tilde{\nu}_i = \tilde{\varepsilon}_{ij} - \tilde{\varepsilon}_{ik}$, and let $\widetilde{H}(x) \equiv \Pr(\widetilde{v}_i \leq x)$. Then, $H(x) = \widetilde{H}(x/\sigma)$. Similarly, we have $h(x) \equiv \dfrac{dH(x)}{dx} = \dfrac{1}{\sigma}\widetilde{h}(x/\sigma)$. Note that

$$h(0) = \frac{1}{\sigma}\widetilde{h}(0).$$

Consider the probability that teacher j wins the contest over student i when her opponent chooses the socially optimal vector of effort, $(\mathbf{e}^*, t^*)$. Let $\mathbf{a} \equiv \mathbf{a}_j = \mathbf{a}_k$. Then this probability is given by

$$\begin{aligned} H(g_i(\mathbf{a}, t_j, \mathbf{e}_j) - g_i(\mathbf{a}, t^*, \mathbf{e}^*)) &= \int_{-\infty}^{g_i(\mathbf{a},t_j,\mathbf{e}_j)-g_i(\mathbf{a},t^*,\mathbf{e}^*)} h(x)dx \\ &= \int_{-\infty}^{g_i(\mathbf{a},t_j,\mathbf{e}_j)-g_i(\mathbf{a},t^*,\mathbf{e}^*)} \frac{1}{\sigma}\widetilde{h}(x/\sigma)dx \end{aligned}$$

Teacher j's problem is given by

$$\max_{\mathbf{e}_j,t_j} NX_0 + (X_1 - X_0)\sum_{i=1}^{N} H(g_i(\mathbf{a}, t_j, \mathbf{e}_j) - g_i(\mathbf{a}, t^*, \mathbf{e}^*)) - C(\mathbf{e}_j, t_j)$$

If we set $X_0 = \frac{U_0 + C(\mathbf{e}^*, t^*)}{N} - \frac{R}{2h(0)}$ and $X_1 - X_0 = R/h(0)$, and use the fact that $\frac{h(x)}{h(0)} = \frac{\widetilde{h}(x/\sigma)}{\widetilde{h}(0)}$ this problem reduces to

$$\max_{\mathbf{e}_j,t_j} R\sum_{i=1}^{N}\left[\int_{-\infty}^{g_i(\mathbf{a},t_j,\mathbf{e}_j)-g_i(\mathbf{a},t^*,\mathbf{e}^*)} \frac{\widetilde{h}(x/\sigma)}{\widetilde{h}(0)}dx\right] + C(\mathbf{e}^*, t^*) - \frac{RN}{2h\,(0)} - C(\mathbf{e}_j, t_j) \quad (3)$$

We first argue that the solution to this problem is bounded in a way that does not depend on σ. Observe that the objective function (**??**) is equal to 0 at $(\mathbf{e}, t) = (\mathbf{e}^*, t^*)$. Any value of $(\mathbf{e}, t)$ that yields a negative value for the objective function thus cannot be optimal. We shall now argue that outside of a compact set, we can be assured that the objective function will generate negative values.

To see this, let us rewrite the first term in (**??**) as follows:

$$\int_{-\infty}^{g_i(\mathbf{a},t_j,\mathbf{e}_j)-g_i(\mathbf{a},t^*,\mathbf{e}^*)} \frac{\widetilde{h}(x/\sigma)}{\widetilde{h}(0)}dx = \int_{-\infty}^{0} \frac{\widetilde{h}(x/\sigma)}{\widetilde{h}(0)}dx + \int_{0}^{g_i(\mathbf{a},t_j,\mathbf{e}_j)-g_i(\mathbf{a},t^*,\mathbf{e}^*)} \frac{\widetilde{h}(x/\sigma)}{\widetilde{h}(0)}dx$$

Since $h(x)$ is symmetric around zero and integrates up to 1, we know the value of the first integral:

$$\int_{-\infty}^{0} \frac{\widetilde{h}(x/\sigma)}{\widetilde{h}(0)} dx = \int_{-\infty}^{0} \frac{h(x)}{h(0)} dx = \frac{1}{2h(0)}$$

The second integral can be bounded. If the upper limit of integration $g_i(\mathbf{a}_j, t_j, \mathbf{e}_j) - g_i(\mathbf{a}, t^*, \mathbf{e}^*)$ is positive, then since $\widetilde{h}(\cdot)$ has a peak at zero, it follows that $\frac{\widetilde{h}(x/\sigma)}{\widetilde{h}(0)} \leq 1$ for all x and so

$$\int_{0}^{g_i(\mathbf{a}_j,t_j,\mathbf{e}_j)-g_i(\mathbf{a},t^*,\mathbf{e}^*)} \frac{\widetilde{h}(x/\sigma)}{\widetilde{h}(0)} dx \leq \int_{0}^{g_i(\mathbf{a}_j,t_j,\mathbf{e}_j)-g_i(\mathbf{a},t^*,\mathbf{e}^*)} dx = g_i(\mathbf{a}_j, t_j, \mathbf{e}_j) - g_i(\mathbf{a}, t^*, \mathbf{e}^*)$$

If instead $g_i(\mathbf{a}_j, t_j, \mathbf{e}_j) - g_i(\mathbf{a}, t^*, \mathbf{e}^*)$ is negative, then since $\widetilde{h}(\cdot) \geq 0$, we have

$$\int_{0}^{g_i(\mathbf{a}_j,t_j,\mathbf{e}_j)-g_i(\mathbf{a},t^*,\mathbf{e}^*)} \frac{\widetilde{h}(x/\sigma)}{\widetilde{h}(0)} dx \leq 0$$

It follows that the objective function for the teacher, (**??**), is bounded above by

$$R \sum_{i=1}^{N} \max\left[g_i(\mathbf{a}, t_j, \mathbf{e}_j) - g_i(\mathbf{a}, t^*, \mathbf{e}^*), 0\right] + C(\mathbf{e}^*, t^*) - C(\mathbf{e}_j, t_j) \tag{4}$$

Now, define the set $U = \left\{ \mathbf{u} \in \mathbb{R}_+^{N+1} : \sum_{i=1}^{N+1} u_i = 1 \right\}$. Any vector $(\mathbf{e}_j, t_j)$ can be uniquely expressed as $\lambda \mathbf{u}$ for some $\lambda \geq 0$ and some $\mathbf{u} \in U$. Given our assumptions on $C(\cdot, \cdot)$, for any vector $\mathbf{u}$ it must be the case that $C(\lambda \mathbf{u})$ is increasing and convex in λ and satisfies the limit $\lim_{\lambda \to \infty} \frac{\partial C(\lambda \mathbf{u})}{\partial \lambda} = \infty$. Since $g_i(\mathbf{a}, \lambda t_j, \lambda \mathbf{e}_j)$ is concave in λ, for any $\mathbf{u} \in U$ there exists a finite cutoff such that the expression in (**??**) evaluated at $(\mathbf{e}_j, t_j) = \lambda \mathbf{u}$ will be negative for all λ above the cutoff. Let $\lambda^*(\mathbf{u})$ denote the smallest cutoff for a given $\mathbf{u}$. Since U is compact, $\lambda^* = \sup\{\lambda^*(\mathbf{u}) : \mathbf{u} \in U\}$ is well defined and finite. It follows that the solution to (**??**) lies in the bounded set $[0, \lambda^*]^{N+1}$.

Next, we argue that there exists a $\overline{\sigma}$ such that for $\sigma > \overline{\sigma}$, the Hessian matrix of second order partial derivatives for this objective function is negative definite over the bounded set $[0, \lambda^*]^{N+1}$. Define $\pi(t_j, \mathbf{e}_j) \equiv R \sum_{i=1}^{N} \left[\int_{-\infty}^{g_i(\mathbf{a},t_j,\mathbf{e}_j)-g_i(\mathbf{a},t^*,\mathbf{e}^*)} \frac{\widetilde{h}(x/\sigma)}{\widetilde{h}(0)} dx \right]$. Then the Hessian matrix is the sum of two matrices, $\mathbf{\Pi} - \mathbf{C}$, where

$$\mathbf{C} \equiv \begin{bmatrix} \frac{\partial^2 C}{\partial e_1^2} & \frac{\partial^2 C}{\partial e_2 \partial e_1} & \cdots & \frac{\partial^2 C}{\partial t \partial e_1} \\ \frac{\partial^2 C}{\partial e_1 \partial e_2} & \frac{\partial^2 C}{\partial e_2^2} & \cdots & \frac{\partial^2 C}{\partial t \partial e_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 C}{\partial e_1 \partial t} & \frac{\partial^2 C}{\partial e_2 \partial t} & \cdots & \frac{\partial^2 C}{\partial t^2} \end{bmatrix}$$

and

$$\mathbf{\Pi} \equiv \begin{bmatrix} \frac{\partial^2 \pi}{\partial e_1^2} & \frac{\partial^2 \pi}{\partial e_2 \partial e_1} & \cdots & \frac{\partial^2 \pi}{\partial t \partial e_1} \\ \frac{\partial^2 \pi}{\partial e_1 \partial e_2} & \frac{\partial^2 \pi}{\partial e_2^2} & \cdots & \frac{\partial^2 \pi}{\partial t \partial e_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \pi}{\partial e_1 \partial t} & \frac{\partial^2 \pi}{\partial e_2 \partial t} & \cdots & \frac{\partial^2 \pi}{\partial t^2} \end{bmatrix}$$

Since the function $C(\cdot)$ is strictly convex, $-\mathbf{C}$ must be a negative definite matrix. Turning to $\mathbf{\Pi}$, since we assume that $H(\cdot)$ is twice differentiable, it follows that $\widetilde{H}(\cdot)$ is also twice differentiable. To determine the limit of $\mathbf{\Pi}$ as $\sigma \to \infty$, let us first evaluate the first derivative of π with respect to e_{ij}:

$$\frac{\partial \pi}{\partial e_{ij}} = \frac{R}{\widetilde{h}(0)} \sum_{m=1}^{N} \left[\frac{\partial g_m(\mathbf{a}_j, t_j, \mathbf{e}_j)}{\partial e_{ij}} \widetilde{h} \left(\frac{g_m(\mathbf{a}_j, t_j, \mathbf{e}_j) - g_m(\mathbf{a}, t^*, \mathbf{e}^*)}{\sigma} \right) \right]$$

Differentiating again yields

$$\frac{\partial^2 \pi}{\partial e_{ij} \partial e_{i'j}} = \frac{R}{\widetilde{h}(0)} \sum_{m=1}^{N} \frac{\partial^2 g_m(\mathbf{a}_j, t_j, \mathbf{e}_j)}{\partial e_{ij} \partial e_{i'j}} \widetilde{h} \left(\frac{g_m(\mathbf{a}_j, t_j, \mathbf{e}_j) - g_m(\mathbf{a}, t^*, \mathbf{e}^*)}{\sigma} \right) + \frac{R}{\sigma \widetilde{h}(0)} \sum_{m=1}^{N} \frac{\partial g_m(\mathbf{a}_j, t_j, \mathbf{e}_j)}{\partial e_{ij}} \frac{\partial g_m(\mathbf{a}_j, t_j, \mathbf{e}_j)}{\partial e_{i'j}} \widetilde{h}' \left(\frac{g_m(\mathbf{a}_j, t_j, \mathbf{e}_j) - g_m(\mathbf{a}, t^*, \mathbf{e}^*)}{\sigma} \right)$$

Because $\frac{\partial g_m(\mathbf{a}_j, t_j, \mathbf{e}_j)}{\partial e_{ij}}$ is finite and $\widetilde{h}'(0) = 0$ given $h(x)$ is maximal at 0, it follows the second term above vanishes in the limit as $\sigma \to \infty$. Hence, we have

$$\frac{\partial^2 \pi}{\partial e_{ij} \partial e_{i'j}} \to \frac{R}{\widetilde{h}(0)} \sum_{m=1}^{N} \left[\frac{\partial^2 g_m(\mathbf{a}_j, t_j, \mathbf{e}_j)}{\partial e_{ij} \partial e_{i'j}} \widetilde{h}(0) \right] = R \sum_{m=1}^{N} \left[\frac{\partial^2 g_m(\mathbf{a}_j, t_j, \mathbf{e}_j)}{\partial e_{ij} \partial e_{i'j}} \right]$$

Given that a similar argument holds for the derivatives of π with respect to t_j, as σ grows, $\mathbf{\Pi}$ converges to an expression proportional to the Hessian matrix for $\sum_{i=1}^{N} g_i$, which is negative semidefinite given each g_i is concave. Hence, the objective function is strictly concave over $[0, \lambda^*]^{N+1}$, the region that contains the global optimum, ensuring the first-order conditions are both necessary and sufficient to define a global maximum.

Here, we have analyzed a two-teacher contest. However, it is straightforward to extend the argument to the case of K rivals. In this case, the solution to teacher j's problem, when all K opponents choose $(\mathbf{e}^*, t^*)$ is the solution to

$$\max_{\mathbf{e}_j, t_j} (X_1 - X_0) \sum_{k=1}^{K} \sum_{i=1}^{N} H(g_i(\mathbf{a}, t_j, \mathbf{e}_j) - g_i(\mathbf{a}, t^*, \mathbf{e}^*)) - C(\mathbf{e}_j, t_j)$$

Thus, since the prizes for the case where there are K opponents are given by $X_1 - X_0 = \frac{R}{Kh(0)}$, this expression reduces to (**??**), and the claim follows immediately.■

Proposition 2: *In the two teacher contest, whenever a pure strategy Nash equilibrium exists, it involves both teachers choosing the socially optimal effort levels* $(\mathbf{e}^*, t^*)$.

Proof of Proposition 2: We begin our proof by establishing the following Lemma:

Lemma: Suppose $C(\cdot)$ is a convex differentiable function which satisfies standard boundary conditions concerning the limits of the marginal costs of each dimension of effort as effort on each dimension goes to 0 or ∞. Then for any positive real numbers $a_1, ..., a_N$ and b, there is a unique solution to the system of equations

$$\begin{aligned} \frac{\partial C(e_1, ..., e_N, t)}{\partial e_i} &= a_i \qquad \text{for } i = 1, ..., N \\ \frac{\partial C(e_1, ..., e_N, t)}{\partial t} &= b \end{aligned}$$

Proof: Define a function $bt + \sum_{i=1}^{N} a_i e_i - C(e_1, .., , e_N, t)$. Since $C(\cdot)$ is strictly convex, this function is strictly concave, and as such has a unique maximum. The boundary conditions, together with the assumption that $a_1, ..., a_N$ and b are positive, ensure that this maximum must be at an interior point. Because the function is strictly concave, this interior maximum and the solution to the above equations is unique, as claimed. ■

Armed with this lemma, we can demonstrate that any pure strategy Nash equilibrium of the two teacher contest involves both teachers choosing the socially optimal effort levels. Note that, given any pure strategy Nash equilibrium, each teacher's choices will satisfy the first order conditions for a best response to the other teacher's actions. Further, since $h(\cdot)$ is symmetric, we know that given the effort choices of j and j',

$$h(\alpha(e_{ij} - e_{ij'}) + t_j - t_{j'}) = h(\alpha(e_{ij'} - e_{ij}) + t_{j'} - t_j)$$

In combination, these observations imply that any Nash equilibrium strategies, $(\mathbf{e}_j, t_j)$ and $(\mathbf{e}_{j'}, t_{j'})$, must satisfy

$$h(0)\frac{\partial C(\mathbf{e}_j, t_j)}{\partial e_{ij}} = R\alpha h(\alpha(e_{ij} - e_{ij'}) + t_j - t_{j'})$$
$$= R\alpha h(\alpha(e_{ij'} - e_{ij}) + t_{j'} - t_j) = h(0)\frac{\partial C(\mathbf{e}_{j'}, t_{j'})}{\partial e_{ij'}}$$

and

$$h(0)\frac{\partial C(\mathbf{e}_j, t_j)}{\partial t_j} = RNh(\alpha(e_{ij} - e_{ij'}) + t_j - t_{j'})$$
$$= RNh(\alpha(e_{ij'} - e_{ij}) + t_{j'} - t_j) = h(0)\frac{\partial C(\mathbf{e}_{j'}, t_{j'})}{\partial t_{j'}}$$

Our lemma implies that these equations cannot be satisfied unless $e_{ij} = e_{ij'} = e^*$ for all $i = 1, ..., N$ and that $t_j = t_{j'} = t^*$. The only pure-strategy equilibrium possible in our two teacher contests is one where teachers invest the classroom instruction effort and common level of tutoring that are socially optimal. ■