



Federal Reserve Bank of Chicago

Pay for Percentile

Gadi Barlevy and Derek Neal

WP 2009-09

PAY FOR PERCENTILE

GADI BARLEVY

FEDERAL RESERVE BANK OF CHICAGO

DEREK NEAL

UNIVERSITY OF CHICAGO AND NBER

Date: August, 2009.

We thank Fernando Alvarez, Julian Betts, Ann Bartel, John Kennan, Kevin Lang, Roger Myerson, Kevin Murphy, Phil Reny, Doug Staiger, Chris Taber, and Azeem Shaikh for helpful comments and discussions. Neal thanks Lindy and Michael Keiser for research support through a gift to the University of Chicago's Committee on Education. Neal also thanks the Searle Freedom Trust. Our views need not reflect those of the Federal Reserve Bank of Chicago or the Federal Reserve System.

ABSTRACT

We propose an incentive pay scheme for educators that links educator compensation to the ranks of their students within appropriately defined comparison sets, and we show that under certain conditions our scheme induces teachers to allocate socially optimal levels of effort to all students. Because this scheme employs only ordinal information, our scheme allows education authorities to employ completely new assessments at each testing date without ever having to equate various assessment forms. This approach removes incentives for teachers to teach to a particular assessment form and eliminates opportunities to influence reward pay by corrupting the equating process or the scales used to report assessment results. Our system links compensation to the outcomes of properly seeded contests rather than cardinal measures of achievement growth. Thus, education authorities can employ our incentive scheme for educators while employing a separate system for measuring growth in student achievement that involves no stakes for educators. This approach does not create direct incentives for educators to take actions that contaminate the measurement of student progress.

1. INTRODUCTION

In modern economies, most wealth is held in the form of human capital, and publicly funded schools play a key role in creating this wealth. Thus, reform proposals that seek to enhance the efficiency of schools are an omnipresent feature of debates concerning public policy and societal welfare. In recent decades, education policy makers have increasingly designed these reform efforts around measures of school output such as test scores rather than measures of school inputs such as computer labs or student-teacher ratios. Although scholars and policy makers still debate the benefits of smaller classes, improved teacher preparation, or improved school facilities, few are willing to measure school quality using only measures of school inputs. During the 1990s many states adopted accountability systems that dictated sanctions and remediation for schools based on how their students performed on standardized assessments. In 2001, the No Child Left Behind Act (NCLB) mandated that all states adopt such systems or risk losing federal funds, and more recently, several states and large districts have introduced incentive pay systems that link the salaries of individual teachers to the performance of their students.

A large empirical literature examines the effects of assessment based incentive systems, but to date, no literature formally explores the optimal design of these systems. Our paper is a first step toward filling this void. We pose the provision of teacher incentives as a mechanism design problem. In our setting, an education authority possesses two sets of test scores for a population of students. The first set of scores provides information about student achievement at the beginning (fall) of a school year. The second set provides information about the achievement of the same population of students at the end (spring) of the school year. Taken together, these test scores provide information concerning the effort that teachers invested in their students.

We begin by noting that if the authority knows the mapping between the test score scale and the expected value of student skill, the authority can implement efficient effort using an incentive scheme that pays teachers for the skills that their efforts helped create. Some will contend that social scientists have no idea how to construct such a mapping and therefore argue that such performance pay systems are infeasible.¹ However, even if policy makers are able to discover the mapping between a particular test score scale and the value of

¹See Balou (2009) for more on difficulties of interpreting psychometric scales. Cawley et al (1999) address the task of using psychometric scales in value-added pay for performance schemes. Cunha and Heckman (2008) describe methods for anchoring psychometric scales to adult outcomes. Their methods cannot be applied to new incentive systems involving new assessments because data on the adult outcomes of test takers cannot be collected before a given generation of students ages into adulthood.

student skill, the authority will find it challenging to maintain this mapping across a series of assessments.

It is well established that scores on a particular assessment become inflated when teachers can coach students on the format or content of specific items. In order to deter this type of coaching, education authorities typically employ a series of assessments that differ in terms of specific item content and form. But, in order to map results from each assessment into a common scale, the authority must equate the various assessment forms, and proper equating often requires common items that link the various forms.² If designers limit the number of common items, they advance the goal of preventing teachers from coaching students for specific questions or question formats, but they hinder the goal of properly equating and thus properly scaling the various assessment forms. In addition, because equating is a complex task and proper equating is difficult to verify, the equating process itself is an obvious target for corruption.³

Given these observations, we turn our attention to mechanisms that require authorities to make incentive payments based only on the ordinal information contained in assessment results, without any knowledge of how the fall and spring assessments are scaled. Because such systems involve no attempt to equate various assessment forms, they can include completely new assessment forms at each point in time and thus eliminate incentives to coach students regarding any particular form of an assessment.

We describe a system called “pay for percentile,” that works as follows. For each student in a school system, first form a comparison set of students against which the student will be compared. Assumptions concerning the nature of instruction dictate exactly how to define this comparison set, but the general idea is to form a set that contains all other students in the system who begin the school year at the same level of baseline achievement in a comparable classroom setting. At the end of the year, give a cumulative assessment to all students. Then, assign each student a percentile score based on his end of year rank *among the students in his comparison set*. For each teacher, sum these within-peer percentile scores over all the students she teaches and denote this sum as a percentile performance index.

²A common alternative approach involves randomly assigning one of several alternate forms to students in a large population, and then developing equating procedures based on the fact that the distribution of achievement in the population receiving each form should be constant. This approach also invites coaching because, at the beginning of the second year of any test based incentive program, educators know that each of their students will receive one of the forms used in the previous period.

³A significant literature on state level proficiency rates under NCLB suggests that political pressures have compromised the meaning of proficiency cutoff scores in numerous states. States can inflate their proficiency rates by making exams easier while holding scoring procedures constant or by introducing a completely new assessment and then producing a crosswalk between the old and new assessment scale that effectively lowers the proficiency threshold. See Cronin et al (2007)

Then, pay each teacher a common base salary plus a bonus that is proportional to her percentile performance index. We demonstrate that this system can elicit efficient effort from all teachers in all classrooms to all students.

The linear relationship between bonus pay and our index does not imply that percentile units are a natural or desirable scale for human capital. Rather, percentiles within comparison sets tell us what fraction of head-to-head contests teachers win when competing against other teachers who educate similar students. For example, a student with a within-comparison set percentile score of .5 performed as well or better than half of his peers. Thus, in our scheme, his teacher gets credit for beating half of the teachers who taught similar students. A linear relationship between total bonus pay and the fraction of contests won works because all of the contests share an important symmetry. Each pits a student against a peer who has the same expected spring achievement when both receive the same instruction and tutoring from their teachers.

The scheme we propose extends the work of Lazear and Rosen (1981). They demonstrate that tournaments can elicit efficient effort from workers when firms are only able to rank the performance of their workers. In their model, workers make one effort choice and compete in one contest. In our model, teachers make multiple effort choices, and these choices may simultaneously affect the outcomes of many contests, but we still find that a common prize for winning each tournament can induce efficient effort. Further, we show that pay for percentile can elicit efficient effort in the presence of heterogeneous gains from instruction, instructional spillovers, and direct peer effects among students in the same classroom. We are able to examine these issues because our model examines contestants who compete in multiple contests simultaneously.

Within our framework, it is natural to think of teachers as workers who perform complex jobs that require multiple tasks because each teacher devotes effort to general classroom instruction as well as one-on-one tutoring for each student. Our results offer new insight for the design of incentives in this setting. Imagine any setting where many workers occupy the same job, and this job requires the execution of multiple tasks. If employers can form an ordinal ranking of worker performance on each task that defines the job, these rankings imply a set of winning percentages for each worker that describe the fraction of workers that perform each task less well than she does. Our main results shows that employers can form an optimal bonus scheme by forming a weighted sum of these winning percentages.

Many engaged in current education policy debates implicitly argue that proper equating of different exam forms is fundamental to sound education policy because education authorities must be able to document the evolution of the distribution of student achievement over

time. But, we argue that education authorities should treat the provision of incentives and the documenting of student progress as separate tasks. Equating studies are not necessary for incentive provision, and the equating process is more likely to be corrupted when high stakes are attached to the exams in question.

Some may worry that our system also provides incentives for teachers to engage in activities that inflate assessment results relative to student subject mastery, and it is true that our system does not address the various forms of outright cheating that test-based incentive systems often invite. However, our goal is to design a system that specifically eliminates incentives for teachers to coach students concerning the answers to particular questions or strategies for taking tests rather than teaching students to master a given curriculum. Further, it is important to note that we are proposing a mechanism designed to induce teachers to teach a given curriculum. We do not address the often voiced concern that potentially important dimensions of student skill, e.g. creativity and curiosity, may not be included in curricular definitions.⁴

2. BASIC MODEL

Here, we describe our basic model and derive optimal teacher effort for our setting. Assume there are J classrooms, indexed by $j \in \{1, 2, \dots, J\}$. Each classroom has one teacher, so j also indexes teachers. We assume all teachers are equally effective in fostering the creation of human capital among their students, and all teachers face the same costs of providing effective instruction. This restriction allows us to focus on the task of eliciting effort from teachers, but it does not allow us to address other issues that arise in settings with heterogeneous teachers, such as how teachers should be screened for hiring and retention and who should be assigned to teach which students.

Each classroom has N students, indexed by $i \in \{1, 2, \dots, N\}$. Let a_{ij} denote the initial human capital of the i -th student in the j -th class. Students within each class are ordered from least to most able, i.e.

$$a_{1j} \leq a_{2j} \leq \dots \leq a_{Nj}$$

We assume all J classes are identical, i.e. $a_{ij} = a_i$ for all $j \in \{1, 2, \dots, J\}$. However, this does not mean that our analysis only applies to an environment where all classes share a common baseline achievement distribution. The task of determining efficient effort for a school system that contains heterogeneous classes can be accomplished by determining

⁴In their well-known paper on multi-tasking, Holmstrom and Milgrom (1991) highlight this concern as an important cost associated with implementing assessment based incentive systems for teachers.

efficient effort for each classroom type. Thus, the planner may solve the allocation problem for the system by solving the problem we analyze for each baseline achievement distribution that exists in one or more classes.⁵

Teachers can undertake two types of efforts to help students acquire additional human capital. They can tutor individual students or teach the class as a whole. Let e_{ij} denote the effort teacher j spends on individual instruction of student i , and t_j denote the effort she spends on classroom teaching. We assume the following educational production function:

$$(2.1) \quad a'_{ij} = g(a_i) + t_j + \alpha e_{ij} + \varepsilon_{ij}$$

The human capital of a student at the end of the period, denoted a'_{ij} , depends on his initial skill level a_i , the efforts of his teacher e_{ij} and t_j , and a shock ε_{ij} that does not depend on teacher effort, e.g. random disruptions to the student's life at home. For now, we assume the production of human capital is separable between the student's initial human capital and all other factors and is linear in teacher efforts. Thus, the tutoring instruction is student-specific, and any effort spent on teaching student i will not directly affect any student $i' \neq i$. Classroom teaching benefits all students in the class. Examples include tasks like lecturing or planning assignments. Here, $g(\cdot)$ is an increasing function and $\alpha > 0$ measures the relative productivity of classroom teaching versus individual instruction, and the productivities of both tutoring effort and classroom instruction are not a function of a student's baseline achievement or the baseline achievement of his classmates. This specification provides a useful starting point because it allows us to present our key results within an analytically simple framework. In section 6, we consider more general production technologies that permit not only heterogeneous gains from instruction among students but also direct peer effects and instructional spillovers within a classroom. The results we derive concerning the efficiency of linking a teacher's bonus pay to her winning percentage in contests against other teachers remain given all of these different generalizations to the production technology, but different technology assumptions may dictate different rules for seeding these contests.

The shocks ε_{ij} are mean zero, pairwise independent for any pair (i, j) , and identically distributed according to a common, continuous distribution $F(x) \equiv \Pr(\varepsilon_{ij} \leq x)$.

⁵We assume the planner takes the composition of each class as given. One could imagine a more general problem where the planner chooses the composition of classrooms and the effort vector for each classroom. However, given the optimal composition of classrooms, the planner still needs to choose the optimal levels of effort in each class. We focus on this second step because we are analyzing the provision of incentives for educators taking as given the sorting of students among schools and classrooms.

Let X_j denote teacher j 's expected income. Then her utility is

$$(2.2) \quad U_j = X_j - C(e_{1j}, \dots, e_{Nj}, t_j)$$

where $C(\cdot)$ denotes the teacher's cost of effort. We assume $C(\cdot)$ is increasing in all of its arguments and is strictly convex. We further assume it is symmetric with respect to individual tutoring efforts, i.e. let \mathbf{e}_j be any vector of tutoring efforts (e_{1j}, \dots, e_{Nj}) for teacher j , and let \mathbf{e}'_j be any permutation of \mathbf{e}_j , then

$$C(\mathbf{e}_j, t_j) = C(\mathbf{e}'_j, t_j)$$

We also impose the usual boundary conditions on marginal costs. The lower and upper limits of the marginal costs with respect to each dimension of effort are 0 and ∞ respectively. These conditions ensure the optimal plan will be interior. Although we do not make it explicit, $C(\cdot)$ also depends on N . Optimal effort decisions will vary with class size, but the tradeoffs between scales economies and congestion externalities at the center of this issue have been explored by others.⁶ Our goal is to analyze the optimal provision of incentives given a fixed class size, N , and here, we suppress reference to N in the cost function.

Let R denote the social value of a unit of a' . Because all students receive the same benefit from classroom instruction, we can normalize units of time such that R can also be interpreted as the gross social return per student when one unit of teacher time is effectively devoted to classroom instruction. Assume that each teacher has an outside option equal to U_0 . An omniscient social planner chooses teacher effort levels in each class $j \in \{1, 2, \dots, J\}$ to maximize the following:

$$\max_{\mathbf{e}_j, t_j} E \left[R \sum_{i=1}^N [g(a_i) + t_j + \alpha e_{ij} + \varepsilon_{ij}] - C(\mathbf{e}_j, t_j) - U_0 \right]$$

Since $C(\cdot)$ is strictly convex, first-order conditions are necessary and sufficient for an optimum. Since all teachers share the same cost of effort, the optimal allocation will dictate the same effort levels in all classrooms, i.e. $e_{ij} = e_i$ and $t_j = t$ for all j . Hence, the optimal effort levels dictated by the social planner, e_1, \dots, e_N and t , will solve the following system of equations:

⁶See Lazear (2001) for example.

$$\frac{\partial C(\mathbf{e}_j, t_j)}{\partial e_{ij}} = R\alpha \quad \text{for } i \in \{1, 2, \dots, N\}$$

$$\frac{\partial C(\mathbf{e}_j, t_j)}{\partial t_j} = RN$$

Given our symmetry and separability assumptions, the cost and returns associated with devoting additional instruction time to a student are not a function of the student's baseline achievement or the distribution of baseline achievement in the class. Thus, in this case, the social optimum dictates the same levels of instruction in all classrooms and the same tutoring effort for all students. Let e^* denote the socially optimal level of individual tutoring effort that is common to all students, t^* denote the efficient level of classroom instruction common to all classes, and (\mathbf{e}^*, t^*) denote the socially optimal effort vector common to all classrooms. In section 6, we generalize the model to allow heterogeneity in returns from instruction, instructional spillovers, and peer effects. In this more general setting, the optimal tutoring effort and classroom instruction for a given student varies with the baseline achievement of the both the student and his classmates. However, the mechanism we propose below still elicits efficient instruction and tutoring for each student in each classroom type.

3. PERFORMANCE PAY WITH INVERTIBLE SCALES

Now consider the effort elicitation problem faced by an education authority that supervises our J teachers. For now, assume that this authority knows everything about the technology of human capital production but cannot observe teacher effort e_{ij} or t_j . Instead, the authority observes test scores that provide a ranking of students according to their achievement at a point in time, $s = m(a)$ and $s' = m(a')$, where $m(a)$ is a strictly monotonic function. For now, assume that this ranking is perfect. Below, we discuss how the possibility of measurement error in the achievement ranks implied by test scores affects our analyses.

Suppose the authority knows $m(\cdot)$, i.e. it knows how to invert the psychometric scale s and recover a . In this setting, there are many schemes that the authority can use to induce teachers to provide socially efficient effort levels. For example, the authority could induce teachers to value improvements in student skill correctly simply by paying bonuses per student equal to Ra'_{ij} . However, from the authority's perspective, this scheme would

be wasteful because it compensates teachers for both the skill created by their efforts and for the stock of skills that students would have enjoyed without instruction, $g(a)$.⁷

If the authority knows both $m(\cdot)$ and $g(\cdot)$, the authority can elicit efficient effort while avoiding this waste by forming an unbiased estimator, V_{ij} , of teacher j 's contribution to student i 's human capital,

$$\begin{aligned} V_{ij} &= a'_{ij} - g(a_{ij}) \\ &= m^{-1}(s'_{ij}) - g(m^{-1}(s_{ij})) \end{aligned}$$

and then paying teachers RV_{ij} per student. Further, even if the authority does not know $g(\cdot)$, it can still provide incentives for teachers based only on their contributions to student skill. For each student i , let the authority form a comparison group composed of all students with the same initial test score as student i at the beginning of the period, i.e. the i -th students from all classrooms. Next, define \bar{a}'_i as the average achievement for this group at the end of the period, i.e.

$$\bar{a}'_i = \frac{1}{J} \sum_{j=1}^J a'_{ij}$$

and consider a bonus schedule that pays each teacher j bonuses linked to the relative performance of her students; specifically $R(a'_{ij} - \bar{a}'_i)$ for each student $i \in \{1, 2, \dots, N\}$. If J is large, teachers will ignore the effect of their choices on \bar{a}'_i , and it is straightforward to show that this bonus scheme elicits efficient effort, (\mathbf{e}^*, t^*) .⁸

Because $\text{plim } \bar{a}'_i = g(a_i) + t^* + \alpha e^*$, the relative achievement of student i , $(a'_{ij} - \bar{a}'_i)$, is not a function of $g(\cdot)$ or a_i in equilibrium. Here, as in the pay for percentile scheme we propose below, teachers receive rewards or penalties for how their students perform relative to comparable students. Moreover, both schemes can be implemented without knowledge of the particular scores associated with each baseline achievement level or the score gains achieved by any student.

⁷Here, we take the assignment of students to classrooms as fixed, and we are assuming that the education authority cannot simply hold an auction and sell teachers the opportunity to earn Ra'_{ij} per student. However, absent such an auction mechanism, we expect any scheme that pays teachers for skills students possess independent of instruction would induce wasteful activities by teachers seeking assignments to high-achieving students.

⁸As an alternative, one can calculate relative performance versus teacher specific means that do not involve the scores of their own students, i.e. $\bar{a}'_{ij} = \frac{1}{J-1} \sum_{k \neq j} a'_{ik}$.

If the authority knows R and $m(\cdot)$, it can implement this bonus scheme using a standard regression model that includes fixed effects for baseline achievement levels and classroom assignment.⁹ Teachers associated with a negative classroom effect will receive a negative bonus or what might be better described as a performance based fine, and total bonus pay over all teachers will be zero. Because expected bonus pay per teacher is zero in this scheme, teachers must receive a base salary that covers their costs. Let X_0 denote the base salary per student. The authority could minimize the cost of eliciting efficient effort by choosing X_0 to satisfy $NX_0 = C(\mathbf{e}^*, t^*) + U_0$.

In this scheme, the focus on variation within comparison sets allows the authority to overcome the fact that it does not know how natural rates of human capital growth, $g(a_i)$, differ among students of different baseline achievement levels, a_i . In the following sections, we demonstrate that by focusing on rank comparisons within comparison sets, the authority can similarly overcome its lack of knowledge concerning how changes in test scores map into changes in human capital at different points on a given psychometric scale.

4. TOURNAMENTS

The scheme described in Section 3 relies on the education authority's ability to translate test scores into the values of students' skills. In order to motivate why the authority might have limited knowledge of how scores map into human capital, suppose the education authority cannot create and administer the assessments but must hire a testing agency to provide \mathbf{s} and \mathbf{s}' , the vector of fall and spring test scores for all students. In order to implement the relative performance scheme we describe above, the authority must announce a mapping between the distribution of student test scores, \mathbf{s}' , and the distribution of reward pay given to the teachers of these students. But, once the authority announces how it will invert $\mathbf{s}' = m(\mathbf{a}')$, it must guard against at least two ways that teachers may attempt to game this incentive system.

To begin, teachers may coach rather than teach, and the existing literature provides much evidence that teachers can inflate student assessment results by giving students the answers to specific questions or having students practice taking tests that contain questions in a specific format.¹⁰ Teachers have opportunity and incentive to engage in these behaviors

⁹For example, if the authority regresses a'_{ij} on only a set of $N + J$ indicator variables that identify baseline achievement groups and classroom assignment, the estimated coefficient on the indicator for teacher j will equal $\frac{1}{N} \sum_{i=1}^N (a'_{ij} - \bar{a}'_i)$, and the authority can multiply these coefficients by RN to determine the total bonus payment for each teacher j .

¹⁰See Jacob (2005), Jacob and Levitt (2002), Klein et al (2000) and Koretz (2002).

whenever the specific items and format of one assessment can be used to predict the items and format present on future assessments.

In order to deter this activity, the education authority could instruct its testing agency to administer exams each fall and spring that cover the same topics but contain different questions in different formats. However, as we noted in the introduction, standard methods do not allow education authorities to place assessments on a common scale if they contain no common items and are given at different points in time.¹¹

Further, taking as given any set of procedures that a testing agency may employ to equate assessment forms, teachers face a strong incentive to lobby the testing agency to alter the content of the spring assessment or its scaling in a manner that weakens effort incentives. Concerns about scale manipulation may seem far fetched to some, but the literature on the implementation of state accountability systems under NCLB contains suggestive evidence that several states inflated the growth in their reported proficiency rates by making assessments easier without making appropriate adjustments to how exams are scored or by introducing new assessments and equating the scales between the old and new assessments in ways that appear generous to the most recent cohorts of students.¹² Teachers facing the relative performance pay scheme described in section 3 above would benefit if they could secretly pressure the testing agency to correctly equate various assessments but then report spring scores that are artificially compressed. If teachers believe their lobbying efforts have been successful, each individual teacher's incentive to provide effort will be reduced, but teachers will still collect the same expected salary. Teachers can achieve a similar result by convincing the testing agency to manipulate the content of the spring exam in a way that compresses scores.

Appendix A fleshes out in more detail the ways that scale dependent systems like the relative performance pay system described above invite coaching and scale manipulation. Given these concerns, we explore the optimal design of teacher incentives in a setting where we require the education authority to employ incentive schemes that are scale invariant, i.e. schemes that rely only on ordinal information and can thus be implemented without regard to scaling. In order to develop intuition for our results, we first consider ordinal contests

¹¹Equating methods based on the assumption that alternate forms are given to populations with the same distribution of achievement are not appropriate for repeated samples taken over time from a given population of students since the hope is that the distribution of true achievement improves over time for a given set of students.

¹²See Peterson and Hess (2006) and Cronin et al (2007). As another example, in 2006, the state of Illinois saw dramatic and incredible increases in proficiency rates that were coincident with the introduction of a new assessment series. See ISBE (2006).

among pairs of teachers. We then examine tournaments that involve simultaneous competition among large numbers of teachers and show that such tournaments are essentially a pay for percentile scheme.

Consider a scheme where each teacher j competes against one other teacher and the results of this contest determine bonus pay for teacher j and her opponent. Teacher j does not know who her opponent will be when she makes her effort choices. She knows only that her opponent will be chosen from the set of other teachers in the system and that her opponent will be facing the same compensation scheme that she faces. Let each teacher receive a base pay of X_0 per student, and at the end of the year, match teacher j with some other teacher j' and pay teacher j a bonus $(X_1 - X_0)$ for each student i whose score is higher than the corresponding student in teacher j' 's class, i.e. if $s'_{ij} \geq s'_{ij'}$. The total compensation for teacher j is thus

$$NX_0 + (X_1 - X_0) \sum_{i=1}^N \mathbb{I}(s'_{ij} \geq s'_{ij'})$$

where $\mathbb{I}(A)$ is an indicator that equals 1 if event A is true and 0 otherwise. Because ordinal comparisons determine all payoffs, teacher behavior and teacher welfare are invariant to any re-scaling of the assessment results that preserves ordering.

For each $i \in 1, 2, \dots, N$, let us define a new variable $\nu_i = \varepsilon_{ij} - \varepsilon_{ij'}$ as the difference in the shock terms for students in the two classes whose initial human capital is a_i . If both teachers j and j' choose the same effort levels, then $s'_{ij} > s'_{ij'}$ if and only if $\nu_i > 0$. Let $H(x) \equiv \Pr(\nu_i \leq x)$ denote the distribution of ν_i . We assume $H(\cdot)$ is twice differentiable and define $h(x) = dH(x)/dx$. Since ε_{ij} are i.i.d, ν_i has mean zero, and $h(\cdot)$ is symmetric around zero. Moreover, $h(x)$ attains its maximum at $x = 0$.¹³

In our framework, ν_i is the only source of uncertainty in our contests, and since test scores rank students perfectly according to their true achievement levels, ν_i only reflects shocks to true achievement. Some readers may wonder how our analyses change if we consider an environment in which test scores rank students according to their true achievement levels with error. Incorporating measurement error makes our notation more cumbersome, but as we now argue, the presence of measurement error in test scores does not alter our basic results if we assume that the errors in measurement are drawn independently from the same distribution for all students.

Suppose test scores are given by $s = m(a + u)$ where u is a random variable with mean zero drawn independently for all students from a common distribution, i.e. each student's

¹³See Vogt (1983).

test score depends on his own human capital, but imperfections in the testing technology create idiosyncratic deviations between the student's true skill and the skill level implied by his test score. In this environment, when both teachers j and j' choose the same effort levels, $s'_{ij} > s'_{ij'}$ if and only if $\nu_i > 0$, where now

$$\nu_i \equiv [g(m^{-1}(s_{ij}) - u_{ij}) + \varepsilon_{ij} + u'_{ij}] - [g(m^{-1}(s_{ij'}) - u_{ij'}) + \varepsilon_{ij'} + u'_{ij'}]$$

As in the case without measurement error in test scores, ν_i is mean zero, and its density is symmetric around zero and maximal at zero. These are the properties of ν_i that we require when proving the results presented below. To simplify our exposition, we proceed under the assumption that test scores provide a perfect ranking of student achievement levels.

Since the initial achievement of the students who are compared to each other is identical, the maximization problem for teacher j is

$$\max_{\mathbf{e}_j, t_j} NX_0 + (X_1 - X_0) \sum_{i=1}^N H(\alpha(e_{ij} - e_{ij'}) + t_j - t_{j'}) - C(\mathbf{e}_j, t_j) - U_0$$

The first order conditions for each teacher are given by

$$(4.1) \quad \frac{\partial C(\mathbf{e}_j, t_j)}{\partial e_{ij}} = \alpha h(\alpha(e_{ij} - e_{ij'}) + t_j - t_{j'})(X_1 - X_0) \text{ for } i = 1, 2, \dots, N$$

$$(4.2) \quad \frac{\partial C(\mathbf{e}_j, t_j)}{\partial t_j} = \sum_{i=1}^N h(\alpha(e_{ij} - e_{ij'}) + t_j - t_{j'})(X_1 - X_0)$$

Consider setting the bonus $X_1 - X_0 = R/h(0)$, and suppose both teachers j and j' choose the same effort levels, i.e. $\mathbf{e}_j = \mathbf{e}_{j'}$ and $t_j = t_{j'}$. Then (4.1) and (4.2) become

$$\begin{aligned} \frac{\partial C(\mathbf{e}_j, t_j)}{\partial e_i} &= R\alpha && \text{for } i \in \{1, 2, \dots, N\} \\ \frac{\partial C(\mathbf{e}_j, t_j)}{\partial t_j} &= RN \end{aligned}$$

Recall that these are the first order conditions for the planner's problem, and thus, the socially optimal effort levels (\mathbf{e}^*, t^*) solve these first order conditions. Nonetheless, the fact that these levels satisfy teacher j 's first order conditions is not enough to show that they are global best responses to the effort decisions of the other teacher. In particular, since $H(\cdot)$ is neither strictly convex nor strictly concave everywhere, the fact that $\mathbf{e}_j = \mathbf{e}^*$ and

$t_j = t^*$ satisfy the first order conditions does not imply that these effort choices are optimal responses to teacher j' choosing the same effort levels.

Appendix B provides proofs for the following two propositions that summarize our main results for two teacher contests:

Proposition 1: *Let $\tilde{\varepsilon}_{ij}$ denote a random variable with a symmetric unimodal density and mean zero, and let $\varepsilon_{ij} = \sigma \tilde{\varepsilon}_{ij}$. There exists $\bar{\sigma}$ such that $\forall \sigma > \bar{\sigma}$, both teachers choosing the socially optimal effort levels (\mathbf{e}^*, t^*) is a pure strategy Nash equilibrium of the two teacher contest.*

The intuition behind the variance restriction in this proposition is straightforward. In any given contest, both effort choices and chance play a role in determining the winner. When chance plays a small role, (\mathbf{e}^*, t^*) will not be a best response to the other teacher choosing (\mathbf{e}^*, t^*) unless prizes are small because, when chance plays a small role, the derivative of the probability of winning of given contest with respect to teacher effort is large. In fact, the bonus, $\frac{R}{h(0)}$, tends to zero as $\sigma \rightarrow 0$. The restriction on σ in Proposition 1 is needed to rule out cases where, given the other teacher is choosing (\mathbf{e}^*, t^*) , the bonus is so small that teacher j 's expected gain from responding with (\mathbf{e}^*, t^*) as opposed to some lower effort level does not cover the incremental effort cost.¹⁴ However, if the element of chance in these contests is important enough, a pure strategy Nash equilibrium exists which involves both teachers choosing the socially optimal effort vectors, (\mathbf{e}^*, t^*) , and Proposition 2 adds that this equilibrium is unique.

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^*) .*

Taken together, our propositions imply that our tournament scheme can elicit efficient effort from teachers who compete against each other in seeded competitions. Thus, the efficiency properties that Lazear and Rosen (1981) derived for a setting in which two players make one effort choice and compete in a single contest carry over to settings in which two players make multiple effort choices and engage in many contests simultaneously. This is true even though some of these effort choices affect the outcome of many contests.

Finally, to ensure that teachers are willing to participate in this scheme, we need to make sure that

$$NX_0 + \frac{RN}{2h(0)} - C(\mathbf{e}^*, t^*) \geq U_0$$

¹⁴Lazear and Rosen (1981) require a similar condition for existence in their single task, two person game.

Given this constraint, the following compensation scheme minimizes the cost of providing efficient incentives

$$\begin{aligned} X_0 &= \frac{U_0 + C(\mathbf{e}^*, t^*)}{N} - \frac{R}{2h(0)} \\ X_1 &= \frac{U_0 + C(\mathbf{e}^*, t^*)}{N} + \frac{R}{2h(0)} \end{aligned}$$

Note that the authority needs only four pieces of information to implement this contest scheme. The authority needs to know each student's teacher, the ranks implied by \mathbf{s} and \mathbf{s}' , and the ratio $\frac{R}{h(0)}$. Recall that R is the gross social return per student generated by one effective unit of classroom instruction. If we stipulate that the authority knows what effective instruction is worth to society but simply cannot observe whether or not effective instruction is being provided, $h(0)$ is the key piece of information that the authority requires.¹⁵

Here, $h(0)$ is the derivative with respect to classroom instruction, t , of the probability that a given teacher wins one of our contests when both teachers are initially choosing the same effort vectors. It will be difficult for any authority to learn $h(0)$ precisely, but one can imagine experiments that could provide considerable information about $h(0)$. The key observation is that, for many different prize levels other than the optimal prize, $\frac{R}{h(0)}$, there exists a symmetric Nash equilibrium among teachers in pure strategies. Thus, given our tournament mechanism and some initial choice for the prize structure, suppose the authority selected a random sample of students from the entire student population and then invited these students to a small number of weekend review classes taught by the authority and not by teachers. If our teachers share a common prior concerning the probability that any one student is selected to participate in these review classes, there will still exist a Nash equilibrium in which both teachers choose the same effort levels. However, given any symmetric equilibrium, the ex post probability that a particular student who received extra instruction will score better than a peer who did not receive extra instruction should increase. Let Δt be the length of the review session. The associated change in the probability of winning is $\Delta p \approx \frac{h(0) + h(\Delta t)}{2} \Delta t$. If we assume that the authority can perfectly monitor

¹⁵In the case of no measurement error in test scores, the assumption that ε_{ij} are distributed identically and independently for all students ensures that $h(0)$ is common to all students. However, with measurement error in test scores and non-linear $g(\cdot)$, $h(0)$ may not be the same for all comparison sets even if both ε_{ij} and u_{ij} are distributed identically and independently for all students. In this case, a variation of our scheme that involves different prizes for contests involving students with different baseline test scores can still elicit efficient effort from all teachers to all students. In section 6, we discuss a related result for the case where the distribution of shocks, ε_{ij} , differs among baseline achievement levels $i = 1, 2..N$.

instruction quality during these experimental sessions and if we choose a Δt that is a trivial intervention relative to the range of shocks, ε , that affect achievement during the year, the sample mean of $\frac{\Delta p}{\Delta t}$ provides a useful approximation for $h(0)$.¹⁶

The two-teacher contest system described here elicits efficient effort from teachers and, because it relies only on ordinal information, can be implemented without equating scores from different assessment forms. Further, since equating is not required, our scheme allows the education authority to employ completely new assessment forms at each point in time and thereby remove opportunities for teachers to coach students for specific questions or question formats based on previous assessments. Our scheme is also robust against efforts to weaken performance incentives by corrupting assessment scales, since the mapping between outcomes and reward pay is scale invariant.

5. PAY FOR PERCENTILE

While our two-teacher contests address some ways that teachers can manipulate incentive pay systems, the fact that each teacher plays against a single opponent may raise concerns about a different type of manipulation. Recall, we assume that the opponent for teacher j is announced at the end of the year after students are tested. Thus, some teachers may respond to this system by lobbying school officials to be paired with a teacher whose students performed poorly. If one tried to avoid these lobbying efforts by announcing the pairs of contestants at the beginning of the year, then one would worry about collusion on low effort levels within pairs of contestants. We now turn to performance contests that involve large numbers of teachers competing anonymously against one another. We expect that collusion on low effort among teachers is less of a concern in this environment.

Suppose that each teacher now competes against K teachers who also have N students. Each teacher knows that K other teachers will be drawn randomly from the population of teachers with similar classrooms to serve as her contestants, but teachers make their effort choices without knowing whom they are competing against. We assume that teachers receive a base salary of X_0 per student and a constant bonus of $(X_1 - X_0)$ for each contest she wins.¹⁷ In this setting, teacher j 's problem is

¹⁶Our production technology implicitly normalizes the units of ε so that shocks to achievement can be thought of in terms of additions to or deletions from the hours of effective classroom instruction t students receive. Further, because R is the social value of a unit of effective instruction time, the prize $\frac{R}{h(0)}$ determined by this procedure is the same regardless of the units used to measure instruction time, e.g. seconds, minutes, hours.

¹⁷A constant prize per contest is not essential for eliciting efficient effort, but we view it as natural given the symmetry of the contests.

$$\max_{\mathbf{e}_j, t_j} NX_0 + \sum_{k=1}^K \sum_{i=1}^N H(\alpha(e_{ij} - e_{ik}) + t_j - t_k)(X_1 - X_0) - C(\mathbf{e}_j, t_j) - U_0$$

The first order conditions are given by

$$(5.1) \quad \frac{\partial C(\mathbf{e}_j, t_j)}{\partial e_{ij}} = \sum_{k=1}^K \alpha h(\alpha(e_{ij} - e_{ik}) + t_j - t_k)(X_1 - X_0) \text{ for } i \in \{1, 2, \dots, N\}$$

$$(5.2) \quad \frac{\partial C(\mathbf{e}_j, t_j)}{\partial t_j} = \sum_{k=1}^K \sum_{i=1}^N h(\alpha(e_{ij} - e_{ik}) + t_j - t_k)(X_1 - X_0)$$

As before, suppose all teachers put in the same effort level, i.e. given any j , $t_j = t_k$ and $\mathbf{e}_j = \mathbf{e}_k$ for $k \in \{1, 2, \dots, K\}$. In this case, the right-hand side of (5.1) and (5.2) reduce to $\alpha Kh(0)(X_1 - X_0)$ and $NKh(0)(X_1 - X_0)$, respectively. Thus, if we set $X_1 - X_0 = \frac{R}{Kh(0)}$ and assume that all teachers choose the socially optimal effort levels, the first order conditions for each teacher are satisfied. Further, Proposition 1 extends trivially to contests among $K > 2$ teachers. Given a similar restriction on the scale parameter σ from Proposition 1 and a prize $\frac{R}{Kh(0)}$ per student, there exists a pure strategy Nash equilibrium in which all teachers choose the socially optimal levels of effort.

Now let $K = (J - 1) \rightarrow \infty$, so that each teacher competes against all other teachers. Further, let A'_i denote a terminal score chosen at random and uniformly from the set of all terminal scores $(a'_{i1}, \dots, a'_{iJ})$. Since the distribution $(a'_{i1}, \dots, a'_{i,j-1}, a'_{i,j+1}, \dots, a'_{iJ})$ converges to the distribution $(a'_{i1}, \dots, a'_{i,j-1}, a'_{ij}, a'_{i,j+1}, \dots, a'_{iJ})$ as $K \rightarrow \infty$, it follows that

$$\lim_{K \rightarrow \infty} \sum_{k=1}^K \frac{\mathbb{I}(a'_{ij} \geq a'_{ik})}{K} = \Pr(a'_{ij} \geq A'_i)$$

and the teacher's maximization problem reduces to

$$\max_{\mathbf{e}_j, t_j} NX_0 + \frac{R}{h(0)} \sum_{i=1}^N \Pr(a'_{ij} \geq A'_i) - C(\mathbf{e}_j, t_j) - U_0$$

This pay for percentile scheme is the limiting case of our simultaneous contests scheme as the number of teachers grows large. Thus, a system that pays teachers bonuses that are proportional to the sum of the within comparison set percentile scores of their students can elicit efficient effort from all teachers.

In our presentation so far, comparison sets contain students who share not only a common baseline achievement level but also by assumption share a common distribution of

baseline achievement among their peers. However, given the separability we impose on the human capital production function in equation (2.1) and the symmetry we impose on the cost function, student i 's comparison set need not be restricted to students with similar classmates. For any given student, we can form a comparison set by choosing all students from other classrooms who have the same baseline achievement level regardless of the distributions of baseline achievement among their classmates. This result holds because the socially optimal allocation of effort (\mathbf{e}^*, t^*) dictates the same level of classroom instruction and tutoring effort from all teachers to all students regardless of the baseline achievement of a given student or the distribution of baseline achievement in his class.

Thus, given the production technology that we have assumed so far, pay for percentile can be implemented quite easily and transparently in any large school system. The education authority can form one comparison set for each distinct level of baseline achievement and then assign within comparison set percentiles based on the end of year assessment results. In the following section, we consider more general production functions. In these environments, comparison sets must condition on classroom characteristics. We show that the existence of peer effects, instructional spillovers and other forces that we have not modeled to this point do not alter the efficiency properties of pay for percentile but simply complicate the task of constructing comparison sets.

6. HETEROGENEOUS GAINS FROM INSTRUCTION AND OTHER GENERALIZATIONS

We now generalize the benchmark model above and show that pay for percentile can be used to elicit socially efficient effort from teachers even when optimal effort for a given student varies with his baseline achievement or is affected by the distribution of baseline achievement among his classmates. Let $\mathbf{a}_j = (a_{1j}, \dots, a_{Nj})$ denote the initial levels of human capital of all students in teacher j 's class, where $j \in \{1, 2, \dots, J\}$. We allow the production of human capital for each student i in class j to depend quite generally on his own baseline achievement, a_{ij} , the composition of baseline achievement within the class, \mathbf{a}_j , the tutoring he receives, e_{ij} , and the tutoring received by all students in his class, \mathbf{e}_j . In the terminology commonly employed in the educational production function literature, we allow heterogeneous rates of learning at different baseline achievement levels given various levels of instruction and tutoring, and we also allow both direct peer effects and instructional spillovers. Formally, the human capital of student i in classroom j is given by

$$(6.1) \quad a'_{ij} = g_i(\mathbf{a}_j, t_j, \mathbf{e}_j) + \varepsilon_{ij}$$

Because $g_i(\cdot, \cdot, \cdot)$ is indexed by i , this formulation allows different students in the same class to benefit differently from the same environmental inputs, i.e. from other classmates, classroom instruction, and individual tutoring of other students in the class. Nonetheless, we place three restrictions on $g_i(\cdot, \cdot, \cdot)$: the first derivatives of $g_i(\cdot, \cdot, \cdot)$ with respect to each dimension of effort are finite everywhere, $g_i(\cdot, \cdot, \cdot)$ is weakly concave, and $g_i(\cdot, \cdot, \cdot)$ depends on class identity, j , only through teacher efforts. Our concavity assumption places restrictions on forms that peer effects and instructional spillovers may take. Our assumption that j enters only through teacher effort choices implies that, for any two classrooms (j, j') with the same composition of baseline achievement, if the two teachers in question choose the same effort levels, i.e. $t_j = t_{j'}$ and $\mathbf{e}_j = \mathbf{e}_{j'}$, the expected human capital for any two students in different classrooms who share the same initial achievement, i.e. $a_{ij} = a_{ij'}$, will be the same. Given this property and the fact that the cost of effort is the same for both teachers, we can form comparison sets at the classroom level and guarantee that all contests are properly seeded.

For now, we will continue to assume the ε_{ij} are pairwise identically distributed across all pairs (i, j) , although we comment below on how our scheme can be modified if the distribution of ε_{ij} can vary across students with different baseline achievement levels. In section 2, given our separable form for $g_i(\cdot, \cdot, \cdot)$, we could interpret the units of ε_{ij} in terms of additions to or deletions from effective classroom instruction time. Given the more general formulation of $g_i(\cdot, \cdot, \cdot)$ here, this interpretation need no longer apply in all classrooms. Thus, the units of ε_{ij} can now only be interpreted as additions to or deletions from the stock of student skill.

We maintain our assumption that the cost of spending time teaching students does not depend on their identity, i.e. $C(\mathbf{e}_j, t_j)$ is symmetric with respect to the elements of \mathbf{e}_j and does not depend on the achievement distribution of the students. Our results would not change if we allowed the cost of effort to depend on the baseline achievement distribution in a class, i.e. $C(\mathbf{a}_j, \mathbf{e}_j, t_j)$, or to be asymmetric with respect to levels of individual tutoring effort, as long as we maintain our assumption that $C(\cdot)$ is strictly convex and is the same for all teachers.

For each class j , the optimal allocation of effort solves

$$(6.2) \quad \max_{\mathbf{e}_j, t_j} \sum_{i=1}^N R[g_i(\mathbf{a}_j, t_j, \mathbf{e}_j) + \varepsilon_{ij}] - C(\mathbf{e}_j, t_j) - U_0$$

Since $g_i(\cdot, \cdot, \cdot)$ is concave for all i and $C(\cdot)$ is strictly convex, this problem is strictly concave, and the first-order conditions are both necessary and sufficient for an optimum. These are given for all j by

$$\begin{aligned}\frac{\partial C(\mathbf{e}_j, t_j)}{\partial e_{ij}} &= R \sum_{m=1}^N \frac{\partial g_m(\mathbf{a}_j, t_j, \mathbf{e}_j)}{\partial e_{ij}} && \text{for } i \in \{1, 2 \dots N\} \\ \frac{\partial C(\mathbf{e}_j, t_j)}{\partial t_j} &= R \sum_{m=1}^N \frac{\partial g_m(\mathbf{a}_j, t_j, \mathbf{e}_j)}{\partial t_j}\end{aligned}$$

For any composition of baseline achievement, there will be a unique (\mathbf{e}_j^*, t_j^*) that solves these equations. However, this vector will differ for classes with different compositions, \mathbf{a}_j , and the tutoring effort, e_{ij} , for each student will generally differ across students in the same class if the students have different initial achievement levels.

We now argue that the same pay for percentile scheme we described above will continue to elicit socially optimal effort vectors from all teachers. The bonus scheme is the same as before, and again, each student will be compared to all students with the same baseline achievement who belong to one of the K other classrooms in his comparison set.¹⁸

Assume that we offer each teacher j a base pay of X_0 per student, and a bonus $X_1 - X_0 = \frac{R}{Kh(0)}$ for each student in any comparison class $k \in \{1, 2 \dots K\}$ who scores below his counterpart in teacher j 's class on the spring assessment. Teacher j 's problem can be expressed as follows:

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

The first order conditions for teacher j are

$$\begin{aligned}\frac{\partial C(\mathbf{e}_j, t_j)}{\partial e_{ij}} &= (X_1 - X_0) \sum_{k=1}^K \sum_{m=1}^N \frac{\partial g_m(\mathbf{a}_j, t_j, \mathbf{e}_j)}{\partial e_{ij}} h(g_m(\mathbf{a}_j, t_j, \mathbf{e}_j) - g_m(\mathbf{a}_k, t_k, \mathbf{e}_k)) \quad \forall i \in \{1, 2 \dots N\} \\ \frac{\partial C(\mathbf{e}_j, t_j)}{\partial t_j} &= (X_1 - X_0) \sum_{k=1}^K \sum_{m=1}^N \frac{\partial g_m(\mathbf{a}_j, t_j, \mathbf{e}_j)}{\partial t_j} h(g_m(\mathbf{a}_j, t_j, \mathbf{e}_j) - g_m(\mathbf{a}_k, t_k, \mathbf{e}_k))\end{aligned}$$

¹⁸As we note at the end of the previous section, this composition restriction on comparison sets is now binding. We noted in the previous section that when $g_i(\cdot, \cdot, \cdot)$ is separable in a_i and teacher's effort, the comparison set for student i may contain any student with the same baseline achievement regardless of the composition of baseline achievement in this student's class.

If all teachers provide the same effort levels, these first order conditions collapse to the planner’s first order conditions. If we assume that other teachers are choosing the socially optimal levels of effort, then for large enough σ , these first-order conditions are necessary and sufficient for an optimal response. The proof of Proposition 1 in Appendix B establishes that there exists a Nash equilibrium such that all teachers choose the first best effort levels in response to a common prize structure in this more general setting as well. However, even though the bonus rate is the same for all teachers, base pay will not be. Because socially efficient effort levels vary with classroom composition, the level of base pay required to satisfy the teachers’ participation constraints will be a function of the specific distribution of baseline achievement that defines a comparison set or a set of competing classrooms.

Here, pay for percentile amounts to competition among teachers within leagues defined by classroom type. These leagues offer properly seeded contests even in the presence of peer effects and heterogeneity in student learning rates. Further, because the competition involves all students in each classroom, teachers internalize the consequences of instructional spillovers.

In practice, it may be impossible to form large comparison sets containing classrooms with identical distributions of baseline achievement. Nevertheless, it may still be possible to implement our system using a large set of quantile regression models that allow researchers to create, for any set of baseline student and classroom characteristics, a set of predicted scores associated with each percentile in the conditional distribution of scores. Given a predicted score distribution for each individual student that conditions on his own baseline achievement and the distribution of baseline achievement among his classmates, education authorities can assign a conditional percentile score to each student and then form percentile performance indices at the classroom level.¹⁹

As we noted above, even with our more general formulation for $g_i(\cdot, \cdot, \cdot)$, the optimal prize structure does not vary with baseline achievement and does not depend on the functional form of $g_i(\cdot, \cdot, \cdot)$. This result hinges on our assumption that the distribution of ε_{ij} , and thus $h(0)$, does not vary among students. When the production function, $g_i(\cdot, \cdot, \cdot)$, varies with achievement level i , experimental estimates of the effect of additional instruction on the probability of winning seeded contests cannot be used to test the hypothesis that $h(0)$ is constant across different baseline achievement levels.²⁰ Further, even if $h(0)$ is the same

¹⁹See Briggs and Betebenner (2009) for an example of how these conditional percentile scores can be calculated in practice.

²⁰In the experiment we described at the end of section 4, students are given a small amount of extra instruction, and the experiment identifies $h(0) \frac{\partial a_{ij}}{\partial t_j}$, which equals $h(0)$ given the linear technology assumed in Section 2. If we ran separate experiments of this type for each baseline achievement level, we could not

for all pairwise student contests regardless of baseline achievement levels, the process of discovering $h(0)$ given the more general technology $g(\cdot, \cdot, \cdot)$ is a little more involved than the process described in section 4. Given this more general technology, R is no longer the value of instruction time provided to any student. R must now be interpreted as the value of the skills created when one unit of instruction time is devoted to a particular type of student in a specific type of classroom. Thus, attempts to learn $h(0)$ based on experiments like those described in section 4 must be restricted to a specific set of students who all have the same baseline achievement level, the same peers, and some known baseline level of instruction and tutoring.

If one assumes that $h(0)$ differs with baseline achievement, the authority can still elicit efficient effort by using a pay for percentile scheme that offers different prizes for winning different contests i.e. rates of pay for percentile, $\frac{R}{Kh_i(0)}$, that are specific to each level of baseline achievement. The key implication of our analyses is that performance pay for educators should be based on ordinal rankings of student outcomes within properly chosen comparison sets. Details concerning the determination of the prize levels associated with different contests hinge on details concerning the nature of education production and shocks to human capital accumulation.

7. LESSONS FOR POLICY MAKERS

In the previous sections, we describe a simultaneous contest mechanism that can elicit efficient effort from teachers and is robust to certain types of manipulation. In this section, we shift our attention to existing performance pay systems and analyze them in light of the lessons learned from our model. Table 1 summarizes a number of existing pay for performance schemes that are currently in operation.

Our model yields several insights that are important but have not been fully recognized in current policy debates. To begin, our analyses highlight the value of peer comparisons as a means of revealing what efficient achievement targets should be. With the exception of TAP and the possible exception of PRP²¹, all the systems described in Table 1 involve setting achievement targets for individual students and rewarding teachers for meeting these

use variation in the product $h_i(0) \frac{\partial a'_{ij}}{\partial t_j}$ to test the hypothesis $h_i(0) = h(0)$ for all i without knowing how $\frac{\partial a'_{ij}}{\partial t_j}$ varies with i .

²¹PRP instructed teachers to document that their students were making progress “as good or better” than their peers. However, PRP involved no system for producing objective relative performance measures, and ex post, the high rate of bonus awards raised questions about the leniency of education officials in setting standards for “as good or better” performance. See Atkinson et al (2009) and Wragg et al (2001).

targets. The targets and rewards in these systems can conceivably be chosen so that teachers respond to such systems by choosing efficient effort. However, the education authority cannot choose these targets correctly without knowing the educational production function, $g_i(\cdot, \cdot, \cdot)$, and the scaling of assessments, $m(\cdot)$. Yet, we noted earlier that the education authority may not observe the production function directly, which raises the possibility that teachers may seek to corrupt the process that the authority uses to determine achievement targets. Further, the scales used by testing agencies to report results to the education authority may also be vulnerable to manipulation.

Table 1
Recent Pay for Performance Systems in Education

Name	Place	Description
ProComp	Denver	Teachers and principals negotiate achievement targets for individual students.
QComp	Minnesota	Schools develop their own plans for measuring teacher contributions to students' achievement.
TAP	14 States	Statistical VAM method produces teacher performance indices.
MAP	Florida	Districts choose their own method for measuring teacher contribution to achievement.
PRP	England	Teachers submit applications for bonus pay and provide documentation of better than average performance in promoting student achievement.

Notes: Each system employs additional measures of teacher performance that are not directly tied to student assessment results. The descriptions presented here describe how performance statistics derived from test scores are calculated.

In contrast, a scheme that pays for performance relative to appropriately chosen peers rather than for performance relative to specific thresholds is less vulnerable to corruption and manipulation. By benchmarking performance relative to peer performance, the education authority avoids the need to forecast $m(g_i(\mathbf{a}_j, \mathbf{e}_j^*, t_j^*) + \varepsilon_{ij})$, i.e. the expected spring score for student i given efficient teacher effort, which it needs to in order to set appropriate thresholds. Moreover, relative performance schemes do not allow teachers to increase total reward pay through influence activities that either lower performance thresholds or inflate the scale used to report assessment results. Total reward pay is fixed in relative

performance systems, and performance thresholds are endogenously determined through peer comparisons.

Among the entries in Table 1, the Value-Added Model (VAM) approach contained in the TAP scheme is the only scheme based on an objective mapping between student test scores and relative performance measures for teachers. Our percentile performance indices provide summary measures of how often the students in a given classroom perform better than comparable students in other schools, while VAM models measure the distance between the average achievement of students in a given classroom and the average achievement one would expect from these students if they were randomly assigned to different classrooms. Both schemes produce relative performance indices for teachers, but the VAM approach is more ambitious. VAM indices not only provide a ranking of classrooms according to performance, they also provide measures of the sizes of performance gaps between classrooms.²²

Cardinal measures of relative performance are a required component of the TAP approach and related approaches to performance pay for teachers because these systems attempt to both measure the contributions of educators to achievement growth and reward teachers according to these contributions. Donald Campbell (1976) famously claimed that government performance statistics are always corrupted when high stakes are attached to them, and the contrast between our percentile performance indices and VAM indices suggests that Campbell's observation may reflect the perils of trying to accomplish two objectives with one set of performance measures. Systems, like TAP, that try to both provide incentives for teachers and produce cardinal measures of educational productivity are likely to do neither well because assessment procedures that enhance an education authority's capacity to measure achievement growth consistently over time introduce opportunities for teachers to game assessment-based incentive systems. Systems that track achievement must place results from a series of assessments on a common scale, and the equating process that creates this common scale will not be credible if each assessment contains a completely new set of questions. The existence of items that have been used on previous assessments and will with some probability be repeated on the current assessment invites teachers to coach students based on the specific items and formats found in the previous assessments, and the existing empirical literature suggests that this type of coaching artificially inflates

²²Some VAM practitioners employ functional form assumptions that allow them to produce universal rankings of teacher performance and make judgements about the relative performance of two teachers even if the baseline achievement distributions in their classes do not overlap. In contrast, our pay for percentile scheme takes seriously the notion that teaching academically disadvantaged students may be a different job than teaching honors classes and provides a context-specific measure of how well teachers are doing the job they actually have.

measures of student achievement growth over time. Further, because it is reasonable to expect heterogeneity in the extent of these coaching activities across classrooms, there is no guarantee that scale dependent measures of relative performance will actually provide the correct ex post performance ranking over classrooms, and as we noted in Section 4, systems that links rewards to cardinal measures of relative achievement growth may introduce political pressures to corrupt equating procedures in ways that compress the distribution of scores, weaken incentives, and contaminate measures of achievement growth.

In contrast, our pay for percentile system elicits effort without creating incentives for coaching or scale manipulation because it permits the use of new assessments at each point in time and contains a mapping between assessment results and teacher pay that is scale invariant. Given our scheme, if education authorities desire measures of secular changes in achievement or achievement growth over time, they can deploy a second assessment system that is scale dependent but with no stakes attached and with only random samples of schools taking these scaled tests. Educators would face no direct incentives to manipulate the results of this second assessment system, and thus by separating the tasks of incentive provision and output measurement, education authorities would likely do both tasks better.

8. CONCLUSION

Designing a set of assessments and statistical procedures that will not only allow policy makers to measure secular achievement growth over time but also isolate the contribution of educators and schools to this growth is a daunting task in the best of circumstances. Further, when the results of this endeavor determine rewards and punishments for teachers and principals, some educators respond by taking actions that artificially inflate measures of student learning. These actions may include coaching students for assessments as well as lobbying testing agencies concerning how results from different assessments are equated. The high stakes testing literature provides much evidence that teachers coach students for high stakes assessments in ways that inflate assessment results relative to student subject mastery, and the literature on NCLB involves significant debate concerning the integrity of proficiency standards. For example, in 2006 in Illinois, the percentage of eighth graders deemed proficient in math under NCLB jumped from 54.3 to 78.2 in one year. This improvement dwarfs gains typically observed in other years and in other states. Because this enormous gain was coincident with the introduction of a new series of assessments that

were scored on a new scale and then equated to previous tests, the entire episode raises suspicions about the comparability of proficiency standards across assessment forms.²³

Our key insight is that properly seeded contests where winners are determined by the rank of student outcomes can provide incentives for efficient teacher effort. Thus, the ordinal content of assessment results provides enough information to elicit socially efficient effort from teachers. This result is important because scale-dependent performance pay systems typically create incentives for educators to coach students based on the questions contained in previous assessments and to pressure testing agencies to alter the scales used to report assessment results. These activities are socially wasteful, and they also contaminate measurements of student achievement. If policy makers desire cardinal measures of how achievement levels are evolving over time or how the contribution of schools to achievement is evolving over time, they will do a better job of providing credible answers to these questions if they address them using a separate measurement system that has no impact on the distribution of rewards and sanctions among teachers and principals.

We are advocating competition based on ranks as the basis for incentive pay systems that are immune to specific corruption activities that plague existing performance pay and accountability systems, but several details concerning how to organize such competition remain for future research. First and foremost, teachers who teach in the same school should not compete against each other. This type of direct competition could undermine useful cooperation among teachers. Further, although we have discussed all our results in terms of competition among individual teachers, education authorities may wish to implement our scheme at the school or school-grade level as a means of providing incentives for effective cooperation.²⁴

In addition, because our scheme is designed to elicit effort from teachers who share the same cost of providing effective effort, it may be desirable to have teachers compete only against other teachers with similar levels of experience, similar levels of support in terms of teacher's aides, and similar access to computers and other resources.²⁵ While more work

²³See ISBE 2006. The new system came online when the state began testing in grades other than grades 3, 5, and 8. The introduction of the new assessment system resulted in significant jumps in both reading and math proficiency in all three grades, but the eighth grade math results are the most suspicious. Cronin et al (2007) contends that other states have inflated proficiency results by compromising the comparability of assessment scales over time.

²⁴This approach is particularly attractive if one believes that peer monitoring within teams is effective. New York City's accountability system currently includes a component that ranks school performance within leagues defined by student characteristics.

²⁵The task of developing a scheme that addresses unobservable differences in teacher talent remains for future research. We have not yet characterized the optimal system for both screening teachers and providing effort incentives based only on the ordinal information in assessments.

remains concerning the ideal means of organizing the contests we describe above, our results demonstrate that education authorities can enjoy important efficiency gains from building incentive pay systems for teachers that are based on the ordinal outcomes of properly seeded contests and that are completely distinct from any assessment systems used to measure the progress of students or secular trends in student achievement. In this paper, our education authority addresses a moral hazard problem given a homogeneous set of teachers. Further research is needed to analyze the potential uses of percentile performance indices in settings where both the education authority and teachers are learning over time about the relative effectiveness of individual teachers.

REFERENCES

- [1] Balou, Dale. "Test Scaling and Value-Added Measurement," NCPI Working paper 2008-23, December, 2008.
- [2] Briggs, Derek and Damian Betebenner, "Is Growth in Student Achievement Scale Dependent," mimeo. April, 2009.
- [3] Campbell, Donald T. "Assessing the Impact of Planned Social Change," Occasional Working Paper 8 (Hanover, N.H.: Dartmouth College, Public Affairs Center, December, 1976).
- [4] Cawley, John, Heckman, James, and Edward Vytlačil, "On Policies to Reward the Value Added of Educators," *The Review of Economics and Statistics* 81:4 (Nov 1999): 720-727.
- [5] Cronin, John, Dahlin, Michael, Adkins, Deborah, and G. Gage Kingsbury. "The Proficiency Illusion." Thomas B. Fordham Institute, October 2007.
- [6] Cunha, Flavio and James Heckman. "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Journal of Human Resources* 43 (Fall 2008): 739-780.
- [7] Gootman, Elissa. "In Brooklyn, Low Grade for a School of Successes." *New York Times*, September 12, 2008.
- [8] Holmstrom, Bengt and Paul Milgrom. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership and Job Design," *Journal of Law, Economics and Organization*, 7 (January 1991) 24-52.
- [9] Illinois State Board of Education. *2006 Illinois State Report Card*, 2006.
- [10] Jacob, Brian. "Accountability Incentives and Behavior: The Impact of High Stakes Testing in the Chicago Public Schools," *Journal of Public Economics* 89:5 (2005), 761-796.
- [11] Jacob, Brian and Steven Levitt. "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating," *Quarterly Journal of Economics* 118:3 (2003) 843-877.
- [12] Klein, Stephen, Hamilton, Laura, McCaffrey, Daniel, and Brian Stecher. "What Do Test Scores in Texas Tell Us?" Rand Issue Paper 202 (2000).
- [13] Koretz, Daniel M. "Limitations in the Use of Achievement Tests as Measures of Educators' Productivity," *Journal of Human Resources* 37:4 (2002) 752-777.
- [14] Lazear, Edward. "Educational Production." *Quarterly Journal of Economics* 16:3 (Aug 2001) 777-803.
- [15] Lazear, Edward and Sherwin Rosen. "Rank Order Tournaments as Optimum Labor Contracts," *Journal of Political Economy* 89:5 (Oct 1981): 841-864.
- [16] Peterson, Paul and Frederick Hess. "Keeping an Eye on State Standards: A Race to the Bottom," *Education Next* (Summer 2006) 28-29.
- [17] Vogt, H. "Unimodality in Differences," *Metrika* 30:1 (Dec 1983) 165-170.

9. APPENDICES

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

$$(9.1) \quad a'_{ij} = t_j + \alpha e_{ij} + \varepsilon_{ij}$$

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 (9.1) 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 productivity of coaching in improving test scores for a student with a given level of 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 $\theta + \mu < 1$ and more generally so that μ is close to zero. 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 $x_j(\mathbf{s}'_1, \dots, \mathbf{s}'_J)$ equal teacher j 's income under the announced rule, where $\mathbf{s}'_j = (s'_{1j}, \dots, s'_{Nj})$ is the vector of scores of all students in class j at the end of the year. We argue in Section 3 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

$$(9.2) \quad x_j(\mathbf{s}'_1, \dots, \mathbf{s}'_J) = NX_0 + R \sum_{i=1}^N \left[s'_{ij} - \frac{1}{J} \sum_{j=1}^J s'_{ij} \right]$$

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

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

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 (9.2) 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 $m(a)$, the expected teacher payoff from (9.2) 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 (9.2), each risk neutral teacher

faces the problem,

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

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 $\hat{\lambda} = 1$. 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 $h(0)$ will differ in these states. However, the procedure we suggest for recovering $h(0)$ works in either state.²⁶

APPENDIX B: PROOFS

Proposition 1: *Let $\tilde{\varepsilon}_{ij}$ denote a random variable with a symmetric unimodal density and mean zero, and let $\varepsilon_{ij} = \sigma \tilde{\varepsilon}_{ij}$. There exists $\bar{\sigma}$ such that $\forall \sigma > \bar{\sigma}$, both teachers choosing the socially optimal effort levels (\mathbf{e}^*, t^*) is a pure strategy Nash equilibrium of the two teacher contest.*

²⁶Note 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.

Proof of Proposition 1: Our proof will be for the general production function $a'_{ij} = g_i(\mathbf{a}, t_j, \mathbf{e}_j)$, which includes the separable case in Section 2 as a special case.

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

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

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 $\mathbf{a} = \mathbf{a}_j = \mathbf{a}_{j'}$. Then this probability is given by

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

Because NX_0 and U_0 are constants, the solution to teacher j 's problem also solves the maximization problem

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

If we set $X_1 - X_0 = R/h(0)$, and use the fact that $\frac{h(x)}{h(0)} = \frac{\tilde{h}(x/\sigma)}{\tilde{h}(0)}$ this problem reduces to

$$(9.3) \quad \max_{\mathbf{e}_j, t_j} R \sum_{i=1}^N \left[\int_{-\infty}^{g_i(\mathbf{a}, t_j, \mathbf{e}_j) - g_i(\mathbf{a}, t^*, \mathbf{e}^*)} \frac{\tilde{h}(x/\sigma)}{\tilde{h}(0)} dx \right] - C(\mathbf{e}_j, t_j)$$

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 (9.3) is nonnegative at $(\mathbf{e}, t) = (\mathbf{0}, 0)$. Next, since $\tilde{h}(\cdot)$ has a peak at zero, it follows that $\frac{\tilde{h}(x/\sigma)}{\tilde{h}(0)} \leq 1$ for all x and so

$$\begin{aligned} \int_{-\infty}^{g_i(\mathbf{a}, t_j, \mathbf{e}_j) - g_i(\mathbf{a}, t^*, \mathbf{e}^*)} \frac{\tilde{h}(x/\sigma)}{\tilde{h}(0)} dx &= \int_{-\infty}^{g_i(\mathbf{a}, 0, \mathbf{0}) - g_i(\mathbf{a}, t^*, \mathbf{e}^*)} \frac{\tilde{h}(x/\sigma)}{\tilde{h}(0)} dx + \int_{g_i(\mathbf{a}, 0, \mathbf{0}) - g_i(\mathbf{a}, t^*, \mathbf{e}^*)}^{g_i(\mathbf{a}_j, t_j, \mathbf{e}_j) - g_i(\mathbf{a}, t^*, \mathbf{e}^*)} \frac{\tilde{h}(x/\sigma)}{\tilde{h}(0)} dx \\ &\leq \int_{-\infty}^{g_i(\mathbf{a}, 0, \mathbf{0}) - g_i(\mathbf{a}, t^*, \mathbf{e}^*)} \frac{\tilde{h}(x/\sigma)}{\tilde{h}(0)} dx + [g_i(\mathbf{a}, t_j, \mathbf{e}_j) - g_i(\mathbf{a}, 0, \mathbf{0})] \end{aligned}$$

The objective function in (9.3) is thus bounded above by

$$(9.4) \quad R \sum_{i=1}^N \int_{-\infty}^{g_i(\mathbf{a}, 0, \mathbf{0}) - g_i(\mathbf{a}, t^*, \mathbf{e}^*)} \frac{\tilde{h}(x/\sigma)}{\tilde{h}(0)} dx + R \sum_{i=1}^N [g_i(\mathbf{a}, t_j, \mathbf{e}_j) - g_i(\mathbf{a}, 0, \mathbf{0})] - C(\mathbf{e}_j, t_j)$$

Note that the first of these sums is bounded above by the finite value RN . Next, define the set $U = \{\mathbf{u} \in \mathbb{R}_+^{N+1} : \sum_{i=1}^{N+1} u_i = 1\}$. 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 $\lim_{\lambda \rightarrow \infty} \frac{\partial C(\lambda \mathbf{u})}{\partial \lambda} = \infty$. Since $R \sum_{i=1}^N [g_i(\mathbf{a}, \lambda \mathbf{u}) - g_i(\mathbf{a}, 0, \mathbf{0})]$ is concave in λ , for any $\mathbf{u} \in U$ there exists a finite cutoff $\lambda^*(\mathbf{u})$ such that the expression in (9.4) evaluated at $(\mathbf{e}_j, t_j) = \lambda \mathbf{u}$ will be negative for all $\lambda > \lambda^*(\mathbf{u})$. Since U is compact, $\lambda^* = \sup\{\lambda^*(\mathbf{u}) : \mathbf{u} \in U\}$ is well defined and finite. It follows that the solution to (9.3) lies in the bounded set $[0, \lambda^*]^{N+1}$ for all σ .

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

differentiable. To determine the limit of $\mathbf{\Pi}$ as $\sigma \rightarrow \infty$, let us first evaluate the first derivative of π with respect to e_{ij} :

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

Differentiating again yields

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

Because $\frac{\partial g_m(\mathbf{a}_j, t_j, \mathbf{e}_j)}{\partial e_{ij}}$ and $\frac{\partial g_m(\mathbf{a}_j, t_j, \mathbf{e}_j)}{\partial e_{i'j}}$ are finite and $\tilde{h}'(0) = 0$, it follows the second term above vanishes in the limit as $\sigma \rightarrow \infty$. Hence, we have

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

Given that a similar argument holds for the derivatives of π with respect to t_j , $\mathbf{\Pi}$ 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 in 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

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

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 (9.3), 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, \dots, a_N and b , there is a unique solution to the system of equations

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

Proof: Define a function $bt + \sum_{i=1}^N a_i e_i - C(e_1, \dots, e_N, t)$. 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, \dots, 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, both 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

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

and

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

Our lemma implies that these equations cannot be satisfied unless $e_{ij} = e_{ij'} = e^*$ for all $i = 1, \dots, N$ 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. ■

Working Paper Series

A series of research studies on regional economic issues relating to the Seventh Federal Reserve District, and on financial and economic topics.

U.S. Corporate and Bank Insolvency Regimes: An Economic Comparison and Evaluation <i>Robert R. Bliss and George G. Kaufman</i>	WP-06-01
Redistribution, Taxes, and the Median Voter <i>Marco Bassetto and Jess Benhabib</i>	WP-06-02
Identification of Search Models with Initial Condition Problems <i>Gadi Barlevy and H. N. Nagaraja</i>	WP-06-03
Tax Riots <i>Marco Bassetto and Christopher Phelan</i>	WP-06-04
The Tradeoff between Mortgage Prepayments and Tax-Deferred Retirement Savings <i>Gene Amromin, Jennifer Huang, and Clemens Sialm</i>	WP-06-05
Why are safeguards needed in a trade agreement? <i>Meredith A. Crowley</i>	WP-06-06
Taxation, Entrepreneurship, and Wealth <i>Marco Cagetti and Mariacristina De Nardi</i>	WP-06-07
A New Social Compact: How University Engagement Can Fuel Innovation <i>Laura Melle, Larry Isaak, and Richard Mattoon</i>	WP-06-08
Mergers and Risk <i>Craig H. Furfine and Richard J. Rosen</i>	WP-06-09
Two Flaws in Business Cycle Accounting <i>Lawrence J. Christiano and Joshua M. Davis</i>	WP-06-10
Do Consumers Choose the Right Credit Contracts? <i>Sumit Agarwal, Souphala Chomsisengphet, Chunlin Liu, and Nicholas S. Souleles</i>	WP-06-11
Chronicles of a Deflation Unforetold <i>François R. Velde</i>	WP-06-12
Female Offenders Use of Social Welfare Programs Before and After Jail and Prison: Does Prison Cause Welfare Dependency? <i>Kristin F. Butcher and Robert J. LaLonde</i>	WP-06-13
Eat or Be Eaten: A Theory of Mergers and Firm Size <i>Gary Gorton, Matthias Kahl, and Richard Rosen</i>	WP-06-14

Working Paper Series *(continued)*

Do Bonds Span Volatility Risk in the U.S. Treasury Market? A Specification Test for Affine Term Structure Models <i>Torben G. Andersen and Luca Benzoni</i>	WP-06-15
Transforming Payment Choices by Doubling Fees on the Illinois Tollway <i>Gene Amromin, Carrie Jankowski, and Richard D. Porter</i>	WP-06-16
How Did the 2003 Dividend Tax Cut Affect Stock Prices? <i>Gene Amromin, Paul Harrison, and Steven Sharpe</i>	WP-06-17
Will Writing and Bequest Motives: Early 20th Century Irish Evidence <i>Leslie McGranahan</i>	WP-06-18
How Professional Forecasters View Shocks to GDP <i>Spencer D. Krane</i>	WP-06-19
Evolving Agglomeration in the U.S. auto supplier industry <i>Thomas Klier and Daniel P. McMillen</i>	WP-06-20
Mortality, Mass-Layoffs, and Career Outcomes: An Analysis using Administrative Data <i>Daniel Sullivan and Till von Wachter</i>	WP-06-21
The Agreement on Subsidies and Countervailing Measures: Tying One's Hand through the WTO. <i>Meredith A. Crowley</i>	WP-06-22
How Did Schooling Laws Improve Long-Term Health and Lower Mortality? <i>Bhashkar Mazumder</i>	WP-06-23
Manufacturing Plants' Use of Temporary Workers: An Analysis Using Census Micro Data <i>Yukako Ono and Daniel Sullivan</i>	WP-06-24
What Can We Learn about Financial Access from U.S. Immigrants? <i>Una Okonkwo Osili and Anna Paulson</i>	WP-06-25
Bank Imputed Interest Rates: Unbiased Estimates of Offered Rates? <i>Evren Ors and Tara Rice</i>	WP-06-26
Welfare Implications of the Transition to High Household Debt <i>Jeffrey R. Campbell and Zvi Hercowitz</i>	WP-06-27
Last-In First-Out Oligopoly Dynamics <i>Jaap H. Abbring and Jeffrey R. Campbell</i>	WP-06-28
Oligopoly Dynamics with Barriers to Entry <i>Jaap H. Abbring and Jeffrey R. Campbell</i>	WP-06-29
Risk Taking and the Quality of Informal Insurance: Gambling and Remittances in Thailand <i>Douglas L. Miller and Anna L. Paulson</i>	WP-07-01

Working Paper Series *(continued)*

Fast Micro and Slow Macro: Can Aggregation Explain the Persistence of Inflation? <i>Filippo Altissimo, Benoît Mojon, and Paolo Zaffaroni</i>	WP-07-02
Assessing a Decade of Interstate Bank Branching <i>Christian Johnson and Tara Rice</i>	WP-07-03
Debit Card and Cash Usage: A Cross-Country Analysis <i>Gene Amromin and Sujit Chakravorti</i>	WP-07-04
The Age of Reason: Financial Decisions Over the Lifecycle <i>Sumit Agarwal, John C. Driscoll, Xavier Gabaix, and David Laibson</i>	WP-07-05
Information Acquisition in Financial Markets: a Correction <i>Gadi Barlevy and Pietro Veronesi</i>	WP-07-06
Monetary Policy, Output Composition and the Great Moderation <i>Benoît Mojon</i>	WP-07-07
Estate Taxation, Entrepreneurship, and Wealth <i>Marco Cagetti and Mariacristina De Nardi</i>	WP-07-08
Conflict of Interest and Certification in the U.S. IPO Market <i>Luca Benzoni and Carola Schenone</i>	WP-07-09
The Reaction of Consumer Spending and Debt to Tax Rebates – Evidence from Consumer Credit Data <i>Sumit Agarwal, Chunlin Liu, and Nicholas S. Souleles</i>	WP-07-10
Portfolio Choice over the Life-Cycle when the Stock and Labor Markets are Cointegrated <i>Luca Benzoni, Pierre Collin-Dufresne, and Robert S. Goldstein</i>	WP-07-11
Nonparametric Analysis of Intergenerational Income Mobility with Application to the United States <i>Debopam Bhattacharya and Bhashkar Mazumder</i>	WP-07-12
How the Credit Channel Works: Differentiating the Bank Lending Channel and the Balance Sheet Channel <i>Lamont K. Black and Richard J. Rosen</i>	WP-07-13
Labor Market Transitions and Self-Employment <i>Ellen R. Rissman</i>	WP-07-14
First-Time Home Buyers and Residential Investment Volatility <i>Jonas D.M. Fisher and Martin Gervais</i>	WP-07-15
Establishments Dynamics and Matching Frictions in Classical Competitive Equilibrium <i>Marcelo Veracierto</i>	WP-07-16
Technology's Edge: The Educational Benefits of Computer-Aided Instruction <i>Lisa Barrow, Lisa Markman, and Cecilia Elena Rouse</i>	WP-07-17

Working Paper Series *(continued)*

The Widow's Offering: Inheritance, Family Structure, and the Charitable Gifts of Women <i>Leslie McGranahan</i>	WP-07-18
Demand Volatility and the Lag between the Growth of Temporary and Permanent Employment <i>Sainan Jin, Yukako Ono, and Qinghua Zhang</i>	WP-07-19
A Conversation with 590 Nascent Entrepreneurs <i>Jeffrey R. Campbell and Mariacristina De Nardi</i>	WP-07-20
Cyclical Dumping and US Antidumping Protection: 1980-2001 <i>Meredith A. Crowley</i>	WP-07-21
Health Capital and the Prenatal Environment: The Effect of Maternal Fasting During Pregnancy <i>Douglas Almond and Bhashkar Mazumder</i>	WP-07-22
The Spending and Debt Response to Minimum Wage Hikes <i>Daniel Aaronson, Sumit Agarwal, and Eric French</i>	WP-07-23
The Impact of Mexican Immigrants on U.S. Wage Structure <i>Maude Toussaint-Comeau</i>	WP-07-24
A Leverage-based Model of Speculative Bubbles <i>Gadi Barlevy</i>	WP-08-01
Displacement, Asymmetric Information and Heterogeneous Human Capital <i>Luojia Hu and Christopher Taber</i>	WP-08-02
BankCaR (Bank Capital-at-Risk): A credit risk model for US commercial bank charge-offs <i>Jon Frye and Eduard Pelz</i>	WP-08-03
Bank Lending, Financing Constraints and SME Investment <i>Santiago Carbó-Valverde, Francisco Rodríguez-Fernández, and Gregory F. Udell</i>	WP-08-04
Global Inflation <i>Matteo Ciccarelli and Benoît Mojon</i>	WP-08-05
Scale and the Origins of Structural Change <i>Francisco J. Buera and Joseph P. Kaboski</i>	WP-08-06
Inventories, Lumpy Trade, and Large Devaluations <i>George Alessandria, Joseph P. Kaboski, and Virgiliu Midrigan</i>	WP-08-07
School Vouchers and Student Achievement: Recent Evidence, Remaining Questions <i>Cecilia Elena Rouse and Lisa Barrow</i>	WP-08-08

Working Paper Series *(continued)*

Does It Pay to Read Your Junk Mail? Evidence of the Effect of Advertising on Home Equity Credit Choices <i>Sumit Agarwal and Brent W. Ambrose</i>	WP-08-09
The Choice between Arm's-Length and Relationship Debt: Evidence from eLoans <i>Sumit Agarwal and Robert Hauswald</i>	WP-08-10
Consumer Choice and Merchant Acceptance of Payment Media <i>Wilko Bolt and Sujit Chakravorti</i>	WP-08-11
Investment Shocks and Business Cycles <i>Alejandro Justiniano, Giorgio E. Primiceri, and Andrea Tambalotti</i>	WP-08-12
New Vehicle Characteristics and the Cost of the Corporate Average Fuel Economy Standard <i>Thomas Klier and Joshua Linn</i>	WP-08-13
Realized Volatility <i>Torben G. Andersen and Luca Benzoni</i>	WP-08-14
Revenue Bubbles and Structural Deficits: What's a state to do? <i>Richard Mattoon and Leslie McGranahan</i>	WP-08-15
The role of lenders in the home price boom <i>Richard J. Rosen</i>	WP-08-16
Bank Crises and Investor Confidence <i>Una Okonkwo Osili and Anna Paulson</i>	WP-08-17
Life Expectancy and Old Age Savings <i>Mariacristina De Nardi, Eric French, and John Bailey Jones</i>	WP-08-18
Remittance Behavior among New U.S. Immigrants <i>Katherine Meckel</i>	WP-08-19
Birth Cohort and the Black-White Achievement Gap: The Roles of Access and Health Soon After Birth <i>Kenneth Y. Chay, Jonathan Guryan, and Bhashkar Mazumder</i>	WP-08-20
Public Investment and Budget Rules for State vs. Local Governments <i>Marco Bassetto</i>	WP-08-21
Why Has Home Ownership Fallen Among the Young? <i>Jonas D.M. Fisher and Martin Gervais</i>	WP-09-01
Why do the Elderly Save? The Role of Medical Expenses <i>Mariacristina De Nardi, Eric French, and John Bailey Jones</i>	WP-09-02
Using Stock Returns to Identify Government Spending Shocks <i>Jonas D.M. Fisher and Ryan Peters</i>	WP-09-03

Working Paper Series *(continued)*

Stochastic Volatility <i>Torben G. Andersen and Luca Benzoni</i>	WP-09-04
The Effect of Disability Insurance Receipt on Labor Supply <i>Eric French and Jae Song</i>	WP-09-05
CEO Overconfidence and Dividend Policy <i>Sanjay Deshmukh, Anand M. Goel, and Keith M. Howe</i>	WP-09-06
Do Financial Counseling Mandates Improve Mortgage Choice and Performance? Evidence from a Legislative Experiment <i>Sumit Agarwal, Gene Amromin, Itzhak Ben-David, Souphala Chomsisengphet, and Douglas D. Evanoff</i>	WP-09-07
Perverse Incentives at the Banks? Evidence from a Natural Experiment <i>Sumit Agarwal and Faye H. Wang</i>	WP-09-08
Pay for Percentile <i>Gadi Barlevy and Derek Neal</i>	WP-09-09