Teacher Incentives

Jaag, Christian

2006

Teacher Incentives

Christian Jaag*

July 6, 2006

Abstract

This paper considers hidden teacher effort in educational production and discusses the implications of multiple teacher effort dimensions on optimum incentive contracts in a theoretical framework. The analysis of educational production in a multitask framework is a new and unique contribution of this paper to the economics of education. We first characterize the first-best and second-best outcomes. The model is extended to address specific questions concerning teacher incentive schemes: We compare input- to output-based accountability measures and study the implication of the level of aggregation in performance measures. Against the background of the empirical evidence on the effectiveness of teacher incentives, we argue that performance measures should be as broad as possible. Further, we present the optimum contract for motivated teachers. Finally, if education is produced in teacher teams, we establish the conditions for optimum team-based and individual incentives: The larger the spillover effects across teacher efforts and the better the measurability of educational achievement, the stronger the case for team-based incentives.

*Institiute of Public Finance and Fiscal Law, University of St.Gallen, Varnbüelstrasse 19, CH-9000 St.Gallen, `christian.jaag@unisg.ch`. Financial support from the Swiss National Science Foundation under project no. 1214-066928 is gratefully acknowledged.

1 Introduction

In this paper, we study optimum teacher incentive schemes. Teachers are often motivated even without external incentives by having subjective preferences as to the flourishing of their pupils. Hence, teacher incentives accounting for this motivation do not primarily aim at increasing their effort but rather at aligning it with the social preferences regarding the goals of education. This section explicitly takes into account multiple effort dimensions in order to illustrate effort substitution which may render merit pay schemes degrading elements in education rather than incentive providers.

There is a large body of literature on optimum incentive schemes, but a considerable gap remains between purely theoretical considerations and empirical evidence about the effectiveness of actual incentives in schools. While theoretical models are usually positive about the effectiveness of incentive mechanisms, it is empirically unclear which mechanisms work and why. This paper aims at filling that gap by an interpretation of the empirical result on grounds of a firm theoretical model. In the context of teaching in multiple effort dimensions, we study the opportunities and drawbacks of incentives in schools with consideration of distorted performance measures, motivated agents, and team efforts.

Educational production is the process in which students acquire skills that make up their educational achievement. It contains all dimensions of learning at school: Students acquire not only technical but also social skills, which enhance their productivity as well as their literacy and promote good citizenship. Educational production is characterized by the following three attributes which distinguish it from production in firms and show that the incentive problems in organizations of the public sector are even aggravated in schools.

Multiple goals The goals of education include in no particular order

- imparting skills of literacy, reasoning and calculation;
- fostering the emotional growth of children;
- preparing students for work by teaching them vocational skills and attitudes suitable for work;
- preparing them for life by teaching them skills of health and financial management;
- preparing them for society by procuring ideals of citizenship and responsibility.

Given the limited resources of schools and teachers, these goals often compete for attention and are therefore substitutes in the production process. However, effort towards certain goals may also affect the achievement of other objectives; e.g. social skills facilitate productive team work among students which itself encourages the emergence of students' appreciation for good citizenship. Our model allows for different degrees of complementarity between various dimensions of the goals of education, which vastly improves the relevance of the model compared to the analysis of only one single dimension of education.

Multiple principals The education system has several collaborators who act as principals in the agency relationship. These include

- parents and children;
- teachers;
- taxpayers;
- potential employers of the graduates;
- society as a whole.

The involved groups in education have diverse preferences and emphases about the multiple goals of education.

Motivated agents Many people enter the teaching profession for idealistic reasons because they enjoy working with children. Introducing incentive schemes may therefore – besides incentivizing certain effort dimensions – destroy teacher's own motivation and have adverse effects on overall teaching. Hence, it is important to note that optimum incentive schemes may not primarily increase teacher efforts, but rather align them with superordinate goals of education.

We start by an overview of the literature on incentives in education in section 2. Section 3 provides a model of educational production by first defining the first-best outcome as a reference case, where the marginal cost of effort just equals the marginal social benefit of the according dimension of education. Then, we discuss the optimum teacher contract (and its limits) if there is only an aggregate performance signal available, which comes from a measure of students' achievement which gages a teacher's *output* in education. Finally, the model is refined and extended in four directions: (1) we consider accountability measures concerning educational *input*, establishing an equivalence result and discussing the conditions for teacher monitoring being the optimum policy; (2) we allow for disaggregate performance measures which extends the space of feasible contracts; (3) we examine the case of motivated teachers, analyzing the optimum response in the incentive contract; (4) we discuss an application of the model to teaching in teams. Section 4 concludes.

In the main sections of the text, we argue mostly intuitively considering special cases; the general formal model supporting the argument and proving the results is provided in the appendix.

2 Related Literature

In our model, multitasking gives rise to distortion if there is only an aggregate performance measure available – even if agents are risk neutral and not protected by limited liability. The seminal contribution to the analysis of moral hazard in multiple dimensions is due to Holmström and Milgrom (1991). They argue that the principal distorts incentives when differences in measurement accuracy lead her to induce the risk averse agent to focus more on some tasks than others. For example, if the tasks are complements at the margin, it is optimal for the principal to reduce the incentives for the task that is easy to measure compared to a situation where the agent is engaged solely in this task. Baker (2002) provides a framework for the analysis of the influence of distortion and risk in a performance measure on their value and use in incentive contracts in an abstract production setting. He argues that the more distorted and the riskier the measure, the less valuable it will be and the less it will be used in an incentive contract. Hence, principals usually face a trade-off between measures that are high risk and low distortion or low risk and high distortion.

The literature discerns incentives systems in schools broadly into the two categories accountability and merit pay, where the establishment of an accountability system is the first step towards the introduction of merit pay. By *accountability* we mean the establishment of some form of standards external to individual educational institutions and the use of tests to assure that teachers or entire schools are doing their best to meet the standards.

The effectivity of monitoring alone in order to incentivize teachers is a controversal issue: Ladd (1999) and Hanushek and Raymond (2004) find positive effects of accountability schemes in schools, while Kane and Staiger (2001), Koretz (2002), and Jacob (2002) are critical, mostly because of the difficulty of designing appropriate accountability measures. Also, the theoretical and empirical relationship between teacher pay and teaching quality is surprisingly controversial:

Hanushek (1994), Eberts, Hollenbeck, and Stone (2002), Lavy (2002, 2003, 2004), and Jürges, Richter, and Schneider (2004) are in favor of teacher merit pay, while Hannaway (1996), Koretz (2002), and Glewwe, Ilias, and Kremer (2003) find that teacher incentives are very hard to implement, and that they tend to crowd out preexisting motivation. Overall, it can be concluded that the concept of individual merit pay measures is theoretically very attractive, while in practice the empirical evidence on its effectiveness is mixed. Potential problems with individual merit pay are: (1) that merit pay may interfere with schools' efforts to promote good teacher performance through pedagogical leadership, encouragement and steps to improve teacher morale; and (2) that it tends to introduce an adversarial atmosphere and create incentives to conceal problems.

A complement to individual merit pay in order to circumvent the above-mentioned problems may be the introduction of merit awards to whole schools. Ladd (1999) studies the experiment with school-based awards in Dallas and finds mixed evidence for a positive effect of such an incentive program on student performance. Awards to whole schools avoid many of the problems of individual merit pay, including the damage to the institutional environment inside the school. However, it introduces the problem of free riding among teachers if social control within a school is weak.

The assumption of purely self-interested individuals may not be appropriate when educational production is considered. In an experimental study, Fehr and Schmidt (2004) find that with explicit monetary incentive, agents indeed concentrate on the tasks they are paid for, so that such incentives need not be optimal. However, with non-binding bonus contracts, concerns for fairness, reciprocity or inequity aversion may affect the principal's bonus payment so that such a contract which builds on trust Pareto dominates the piece-rate contract. In schools, it seems reasonable that the interaction between teachers and students rests upon reciprocity. In the relationship between teachers and an anonymous school authority, which is discussed in the following, this aspect is surely less important.

3 The Model

3.1 Model Outline and Reference Case

The education process involves the sequence of events as displayed in figure 1.

Figure 1

The sequence of decisions in the education model.

1 Government sets teacher pay schedule, and class size;

2 Teacher chooses effort;

3 Nature resolves risk.

In the first stage, the government fixes its education policy which is fully characterized by prescribed class size, teacher remuneration schedule and possibly monitoring activities to learn about teacher effort. Subsequently, teachers decide on their effort which is – together with class size – a determinant of their students' success probability. In the third stage, nature resolves risk and the government pays teachers according to their wage schedule.

We assume that there exists some aggregate measure h of what a student has been endowed with per time-unit of effective schooling. Some aspects of h can be well assessed in tests, e.g. science skills, which are often referred to as *hard skills*. Other aspects are much harder to evaluate, e.g. social skills and a other virtues such as a student's constructive attitude towards society, which are called *soft skills*. There are n different skills which are produced by an according specific teacher effort and contribute to educational achievement separably: $h(\mathbf{e}) = \sum_{i=1}^{n} \alpha_i e_i$.[1] Hence, a teacher's performance translates directly to her students' performance. The marginal productivity α_i can be interpreted as the marginal value society assigns to skills in dimension i. We model educational production as a function of class size m and a number of dimensions of teacher effort. The time during which teaching is effective decreases in the number of students in a class, such that student achievement is given by

$$P(m, \mathbf{e}) = \pi^m h = \pi^m \sum_i \alpha_i e_i. \tag{1}$$

The parameter π denotes the probability that a student does not disturb classwork during any moment in time; effective teaching takes place only if nobody disrupts, which is the case in a fraction π^m of time spent in class. Associated with teacher efforts is a quadratic effort cost function of the form

$$C(\mathbf{e}) = \sum_{i=1}^{n} \sum_{j=1}^{n} e_i c_{ij} e_j, \tag{2}$$

[1] Bold variables denote vectors, e.g. $\mathbf{e}$ is an array of a number of one-dimensional e_i.

which allows various efforts to interact. We assume that efforts complement each other such that existing effort in one dimension lowers the marginal cost of exerting effort in an other dimension ($c_{ij} < 0 \; \forall i \neq j$):

$$\frac{\partial C}{\partial e_j} = \sum_i e_i c_{ij},$$
$$\frac{\partial}{\partial e_i}\left(\frac{\partial C}{\partial e_j}\right) = c_{ij} < 0.$$

Our measure of educational production P comprises soft and hard skills, which in our model differ simply by their measurability via tests or school monitoring. Of course, hard skills are also more easily observed on the labor market and are thus more likely to be compensated. Soft skills often give rise to external effects: They may reduce stealing, corruption and freeriding in teamwork or increase political participation with important returns to society. Whether education is to the students' private benefit or beneficial to society is irrelevant in our model, though: The school authority is assumed to maximize total social welfare, irrespective of where it accrues and with an arbitrary weighting of various skill dimensions.

For the sake of a reference, we first assume that teacher effort in every dimension is directly observable such that it can be contracted upon in order to achieve the socially desired mix of skills – hence, there is no need to rely on a distorted signal. We restrict the analysis in the main part of the paper to two dimensions while the formal treatment in the appendix allows for an arbitrary number of effort dimensions.

The first-best allocation under full information is characterized by the following result.

Result 1 *(a) Optimum class size increases in the probability that students behave well but is independent of the other factors in educational production. (b) Optimum efforts increase in marginal productivity and class size and decrease in marginal cost.*

Proof. The result corresponds to equations (5) and (6) in the appendix. ∎

In order to achieve the optimum overall distortion in class, it must be the case that better behaved students be taught in larger classes. Result 1b follows directly from the first-order condition that marginal benefits equal marginal costs. In addition to result 1, total social welfare is stated in the appendix as reference for allocations resulting from various incentive mechanisms under asymmetric information.

Example 2 *Consider the case of two effort dimensions with the simplification that effort costs are independent across effort dimensions:* $c_{12} = c_{21} = 0$. *Education is produced according to* $P(m, e_1, e_2) = \pi^m (\alpha_1 e_1 + \alpha_2 e_2)$ *and the cost function per student writes as* $\frac{1}{m} C(e_1, e_2) = \frac{1}{m}\left(e_1^2 c_{11} + e_2^2 c_{22}\right)$. *Optimum efforts are found by solving* $W^* = \max_{m, e_1, e_2 \in \mathbb{R}_+} \left\{ P(m, e_1, e_2) - \frac{1}{m} C(e_1, e_2) \right\}$. *The first-order conditions (equating marginal benefits and marginal costs) with respect to* e_1, e_2 *and* m *are*

$$\begin{aligned}
\pi^m \alpha_1 &= \frac{2}{m} e_1 c_{11}, \\
\pi^m \alpha_2 &= \frac{2}{m} e_2 c_{22}, \\
\pi^m \ln \pi \left(\alpha_1 e_1 + \alpha_2 e_2\right) &= -\frac{1}{m^2} \left(e_1^2 c_{11} + e_2^2 c_{22}\right),
\end{aligned}$$

and solving for e_1, e_2 *and* m *yields*

$$\begin{aligned}
e_1^* &= \frac{1}{2} m^* \pi^{m^*} \frac{\alpha_1}{c_{11}}, \\
e_2^* &= \frac{1}{2} m^* \pi^{m^*} \frac{\alpha_2}{c_{22}}, \\
m^* &= -\frac{1}{2 \ln \pi}.
\end{aligned}$$

Total welfare is given by

$$W^* = -\frac{1}{8 \ln(\pi) \epsilon} \left(\frac{\alpha_1^2}{c_{11}} + \frac{\alpha_2^2}{c_{22}} \right),$$

where ϵ *denotes Euler's number. Note that since effort is assumed to be observable, there is no need to draw on student performance tests by which various effort dimensions would be better or worse assessable.*

The basic model considers a situation in which teacher effort is not observable to the school authority, but an aggregate signal representing student performance is. By the term *signal* we refer to an observable performance measure of a teacher and her students. An aggregate signal is e.g. the overall score in a standardized test, as in the Program for International Student Assessment (PISA). A disaggregate signal is a more specific score, e.g. in a reading or math test.

3.2 Basic Model With Unobservable Teacher Effort

If effort is not directly observable, there is a principal-agent relationship between a school authority which intends to maximize the net surplus from education as principal and teachers as agents who are interested in maximizing their net benefit from teaching. The measurement of the students' accomplishments in different skill dimensions has to be based on a signal which is also produced in the education process and which is – as opposed to actual performance – actually measurable: $S\,(m,\mathbf{e}) = \pi^m \sum_i a_i(e_i+\varepsilon_i)$. a_i is the marginal productivity of effort e_i in the signal's dimension i and ε_i is the associated observation error. The vector $(\varepsilon_1,\ldots\varepsilon_n)'$ is distributed joint-normal $N_n\,(0,\mathbf{\Sigma})$. The difference between α_i and a_i and hence the *distortion* in the performance measure is due to the fact that activity i affect the teacher's objective differently from how it affects the performance measure. There is also an error term ε_i which contaminates the signal, accounting for influences on test results which cannot be influenced by the teacher. We call this the noise in the performance measure. For the sake of simplicity, we assume that the signal is homogeneous for students in the same class.

The school authority designs an optimum compensation schedule $T(S(m,\mathbf{e})) = b + tS(m,\mathbf{e})$ for teachers, where b is base salary and t denotes the slope of the pay schedule (*power* of the incentive contract).[2] Since unobservable effort is not directly contractible upon, teachers maximize their own utility:

$$U(x) = -exp(-xr) \tag{3}$$

with respect to their efforts, where their net benefit is $x = T(S(m,\mathbf{e})) - C(\mathbf{e})$ and r denotes the Arrow-Pratt coefficient of absolute risk aversion. The optimum effort choice is the incentive constraint (IC) to the authority's welfare maximization problem. In addition, there is a participation constraint (PC) guaranteeing that teachers receive at least the utility level of their outside opportunity and hence are willing to enter the profession. For simplicity, we assume the outside opportunity to equal zero, but in the presence of free government funds it can assume any value without fundamentally altering the results.[3]

[2]For the sake of simplicity, we follow the literature in restricting ourselves to linear contracts. A fully flexible compensation scheme $T(S)$ may not be feasible in Europe, where most teachers are civil servants. However, many countries have recently introduced teacher evaluation systems with financial incentives. In the Canton of Zürich, Switzerland, such a system has been introduced in 1999 (Lohnwirksames Qualifikations-System, LQS).

[3]In the case of costly government resources, the case is more involved, cf. Jaag (2005).

With two effort dimensions, the linear form of the educational production function implies that society weighs both dimensions with α_1 and α_2, respectively. The resulting optimum allocation of efforts could be matched by an appropriate emphasis on various efforts, i.e. by compensating various performance dimensions differently. However, available performance measures may deviate from actual performance. Consider an incentive program for school teachers that uses student test scores as the performance measure. The true objective of education be good citizenship and successful scientists. There are many things that a teacher can do to achieve these objectives, some of which might also improve according test scores. A teacher also can do things that will improve the performance measure while having little effect on the true objectives of the school system (*teaching to the test*). Knowing how teacher effort affects both actual and measurable student performance via coefficient vector **ff** $= (\alpha_1, \ldots \alpha_n)'$ and $\mathbf{a} = (a_1, \ldots a_n)'$, respectively, one can infer and compensate the actual teacher effort from test measures. Let e_1 enhance the pupils' soft skills while e_2 is the effort put into to teaching of hard skills. This situation is depicted in figure 2. In the graph, **a** and **ff** indicate the relative signal strength and actual performance in dimension 1 and 2 per unit of e_1 and e_2 respectively. Effort in the first dimension, e_1, greatly adds to productivity, but only little to the signal, while effort e_2 adds more to the signal than to actual productivity. Hence, performance pay based on the signal without accounting for its bias distorts incentives. If actual productivity is orthogonal to the signals which are generated by the same efforts, the signal contain no information, such that paying performance becomes futile. In figure 2, this would be the case if the angle θ between **ff** and **a** were equal to 90°.

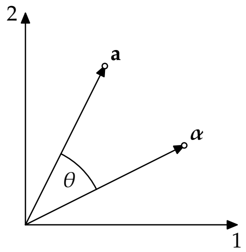

Figure 2

Geometrical interpretation of the signal distortion in two effort dimensions.

Result 3 *With only an aggregate performance measure (signal) available, the optimum power of incentives decreases in the distortion and the noise in the performance measure* *as well as in the teacher's degree of risk aversion; it increases in the probability students behave well.*

Proof. The results correspond to (13) in the appendix. ■

This result implies that in riskier environments, there should be lower powered incentives in favor of larger base salaries because there is a trade-off of risk and incentives. Prendergast (2000) argues that, in general, there is no such trade-off in actual incentive contracts, observing that indeed much of the use of incentive pay is in volatile industries. However, in a school setting, the trade-off is apparent in the common use of well measurable science tests, while the students' citizenship hardly enters a teacher's pay schedule. When test scores do not coincide with society's objectives, such performance measures will induce teachers to engage in dysfunctional behavior which increases the performance measure possibly without increasing the school's real objective. This could result in a performance measure uncorrelated with what the school system cares about.

Comparing the two welfare measures (7) and (14) in the appendix, we see that asymmetric information clearly reduces welfare due to the distortion and the noise of the performance measure, but also due to the teacher's risk bearing.

The school authority maximizes total educational achievement minus its transfer to the educator which consists of a base salary plus a signal-dependent performance pay. The two conditions restricting the authority are the teacher's participation constraint which guarantees that a teacher may leave the profession if she has a better outside option and an incentive constraint which takes into account that the teacher – given the contract – weighs her efforts optimally (from an individual point of view). The availability of only an aggregate signal does not allow for effort-specific incentives. Hence, the school authority as the principal faces the trade-off between excessively incentivizing science teaching and underweighing the procurement of social competence. If teachers are risk-averse, there is an additional trade-off between insurance and efficiency: If teachers are held fully liable for their students' performance, they behave efficiently but bear all the risk, if they are paid a flat salary, they are fully insured, but have no incentive to exert the optimum amounts of effort.

Example 4 *(1 cont.) We continue the example of the previous section, additionally assuming that signal distortions are independent across effort dimensions, such that the off-diagonal entries in the covariance matrix* $\boldsymbol{\Sigma}$ *are equal to zero,* $\sigma_{12} = \sigma_{21} = 0$. σ_1^2 *and*

σ_2^2 denote the variance in the observation error in dimension 1 and 2 respectively. The education authority now maximizes the social net benefit of education as above, where now the cost of the teacher's per-student risk premium $\frac{rt^2\pi^{2m}}{2m}\left(a_1^2\sigma_1^2+a_2^2\sigma_2^2\right)$ adds to effort costs:[4]

$$W_A^{SB}=\max_{m\in\mathbb{R}_+,t\in\mathbb{R}}\left\{P\left(m,e_1,e_2\right)-\frac{1}{m}C\left(e_1,e_2\right)-\frac{rt^2\pi^{2m}}{2m}\left(a_1^2\sigma_1^2+a_2^2\sigma_2^2\right)\right\},$$

where e_1, e_2 are the efforts chosen by the teacher maximizes her benefit from teaching minus effort costs, and who therefore solves

$$\{e_1,e_2\}\in\arg\max_{\{\tilde{e}_1,\tilde{e}_2\}\in\mathbb{R}_+^2}\left\{t\pi^m\left(a_1\tilde{e}_1+a_2\tilde{e}_2\right)-\left(\tilde{e}_1^2c_{11}+\tilde{e}_2^2c_{22}\right)\right\}.$$

The first-order conditions equate marginal benefits and marginal costs:

$$\begin{aligned}t\pi^m a_1&=2e_1c_{11},\\ t\pi^m a_2&=2e_2c_{22}.\end{aligned}$$

Solving for e_1, e_2 yields

$$\begin{aligned}e_1&=\frac{t}{2}\pi^m\frac{a_1}{c_{11}},\\ e_2&=\frac{t}{2}\pi^m\frac{a_2}{c_{22}}.\end{aligned}$$

Substituting these values into the objective function yields the optimum slope of the incentive contract

$$t_A^{SB}=m_A^{SB}\frac{\frac{\alpha_1a_1}{c_{11}}+\frac{\alpha_2a_2}{c_{22}}}{\frac{a_1^2}{c_{11}}+\frac{a_2^2}{c_{22}}+2r\left(a_1^2\sigma_1^2+a_2^2\sigma_2^2\right)}$$

and optimum class size $m_A^{SB}=m^*$ such that efforts write as

$$\begin{aligned}e_{1,A}^{SB}&=e_1^*\frac{\frac{\alpha_1a_1}{c_{11}}+\frac{\alpha_2a_2}{c_{22}}}{\frac{a_1^2}{c_{11}}+\frac{a_2^2}{c_{22}}+2r\left(a_1^2\sigma_1^2+a_2^2\sigma_2^2\right)},\\ e_{2,A}^{SB}&=e_2^*\frac{\frac{\alpha_1a_1}{c_{11}}+\frac{\alpha_2a_2}{c_{22}}}{\frac{a_1^2}{c_{11}}+\frac{a_2^2}{c_{22}}+2r\left(a_1^2\sigma_1^2+a_2^2\sigma_2^2\right)}.\end{aligned}$$

[4] With exponential utility, as in (3), $EU(x)=-\int exp(-rx)f(x)dx=-exp\left(-r\left[Ex-0.5r\sigma_x^2\right]\right)$. Note that expected utility is increasing in $Ex-0.5r\sigma_x^2$, which means that we can take a monotonic transformation of expected utility and use the utility function $U\left(Ex,\sigma_x^2\right)=Ex-0.5r\sigma_x^2$.

Note that the power of the incentive contract depends on the noise and distortion of both effort dimensions. Since only an aggregate signal is available, incentives cannot be dimension-specific, such that the distortion is the same in both effort dimensions.

We can interpret the result as follows: Equilibrium second-best efforts are first-best if the signal is not distorted $(\alpha_i = a_i,\ i \in \{1,2\})$ *and the teacher is risk-neutral* $(r = 0)$ *or there is no risk* $\left(\sigma_i^2 = 0\right)$. *They depart from the first-best allocation due to the cost associated with the distortion of the signal if* $\alpha_i \neq a_i$ *and/or the cost of the teacher's risk premium if* $r \neq 0$ *or* $\sigma_i^2 \neq 0$. *The signal distortion demands for an excessively high power of the incentive contract, while a risk averse teacher needs to be compensated for her risk-taking. Total welfare writes as*

$$W_A^{SB} = -\frac{1}{8\ln(\pi)\,\epsilon}\frac{\left(\frac{\alpha_1 a_1}{c_{11}} + \frac{\alpha_2 a_2}{c_{22}}\right)^2}{\frac{a_1^2}{c_{11}} + \frac{a_2^2}{c_{22}} + 2r\left(a_1^2\sigma_1^2 + a_2^2\sigma_2^2\right)}.$$

It decreases in the noise and the distortion of the performance measure and the teacher's risk aversion. If e.g. actual performance in one skill dimension is not measurable at all $(\sigma \to \infty)$ *teachers cannot be incentivized and total welfare drops to zero since a flat-salary, which would be optimal in this case, does not incentivize the teacher at all. However, in reality, this will not happen, of course, since teachers are usually self-motivated and maintain certain effort levels without according incentives (cf. section 3.5 below).*

3.3 Input-based Signal

In the previous section, we have assumed that the teacher's performance is measured via tests on students. These test measures were not only distorted, but also affected by class size and the students' behavior in class. Likewise, one could also measure a teacher's performance directly and possibly more efficiently before her interaction with students. This could be achieved by not testing students, but visiting class and observing teaching directly. The relevant signal then becomes $S(m, \mathbf{e}) = \sum_i a_i (e_i + \varepsilon_i)$. The potential benefit of an input-based performance measure comes from the direct avoidance of the class size terms in the teacher's utility maximization since she receives her pay irrespective of class size. Principally, the compensation of her income risk can thus be smaller than with an output-based signal.

Result 5 *Measuring teacher performance directly at the input does not increase the overall efficiency of the incentive contract.*

Proof. The result follows from a comparison of equations (14) and (15) in the appendix. ■

An even stronger input-based performance measure is perfect teacher monitoring such that the signal $S(\mathbf{e}) = \mathbf{a}'\mathbf{e}$ is available. This may be the case if there are school inspectors who monitor class work closely or if teachers are urged to account for their teaching by supplying extensive reports about their work to the school authority. We assume that this is only possible at an additional monitoring cost $\mu C(\mathbf{e})$.[5] While merit pay is a priori costless to society, since it consists of a pure transfer from the government to teachers, monitoring appears to be socially wasteful since it is not productive in any sense – other than its contribution to uncover possibly hidden teacher effort. Hence, there seems to be a strong case in favor of merit pay. However, practical experience shows that merit pay is very rarely employed as teacher incentive program. This is in fact optimal in the case that the cost of teacher information rent outweighs the cost associated with monitoring. Consequentially, the school authority prefers monitoring over merit pay.

Result 6 *The attractiveness of a-priori wasteful monitoring is the higher the lower its cost, the higher the teacher's degree of risk-aversion and the larger the variance in the performance measure.*

Proof. The result follows from (17) in the appendix. ■

This extension allows to understand why incentive contracts are rarely employed in schools: There is a social cost attached to incentivize teachers, such that other forms of stimulation may be more effective (cf. Jaag, 2005).

Example 7 *(1 cont.) Compare the result in example 4 with the allocation which results from the solution of the problem* $\max\limits_{m,t\in\mathbb{R}_+}\left\{P\left(m,e_1,e_2\right)-\frac{1}{m}\left(1+\mu\right)C\left(e_1,e_2\right)\right\}$ *where* e_1 *and* e_2 *are chosen by the teacher according to her incentive contract. Total surplus with monitoring is*

$$W_{NL}^{SB}=-\frac{1}{8\ln\left(\pi\right)\epsilon}\frac{\left(\frac{\alpha_1 a_1}{c_{11}}+\frac{\alpha_2 a_2}{c_{22}}\right)^2}{\left(1+\mu\right)\left(\frac{a_1^2}{c_{11}}+\frac{a_2^2}{c_{22}}\right)}.$$

[5]The proportionality to effort cost is assumed for computational simplicity. It can be argued that the first units of effort are easily observable in class, while higher efforts concern preparation work which is far more difficult to observe. Hence, the assumption of progressive monitoring costs is plausible.

Hence, direct teacher monitoring dominates the incentive contract iff

$$\mu < 2r \frac{a_1^2\sigma_1^2 + a_2^2\sigma_2^2}{\frac{a_1^2}{c_{11}} + \frac{a_2^2}{c_{22}}}.$$

3.4 Disaggregate Signal

The above-mentioned trade-off between overweighing one dimension and underweighing the other can be avoided when disaggregated signals are specific to each separate dimension. Since also the power of incentives can be chosen dimension-specifically, there is no longer a distortion due to the multitasking nature of educational production.

Result 8 *If a disaggregate performance measure is available, optimum performance pay emphasizes performance dimensions with low noise and distortion.*

Proof. The result corresponds to equation (20) in the appendix. ∎

Using disaggregate signals, total welfare can be substantially increased because of the feasibility of dimension-specific fine-tuning of the power of incentives. This explains why performance pay systems with differentiated measures, such as in Dallas (cf. Ladd, 1999), work well while others based on only a few performance dimensions fail due to effort substitution: With single-dimension effort measures, incentives are either too weak or degrading, while disaggregate performance measures allow for the optimum power in each dimension.

Example 9 *(1 cont.) In the case that the two signals s_1 and s_2 are observable separately, the power of incentives in the two dimensions can be adjusted accordingly. Hence, the principal's problem writes as $\max_{m,t\in\mathbb{R}_+} \left\{ P(m,e_1,e_2) - \frac{1}{m}C(e_1,e_2) - \frac{r\pi^{2m}}{2m}\left(a_1^2\sigma_1^2t_1^2 + a_2^2\sigma_2^2t_2^2\right) \right\}$ where again e_1 and e_2 are chosen by the teacher according to her incentive contract. The optimum power of incentives now depends on the properties of various effort dimensions:*

$$t_{1,D}^{SB} = \frac{m_D^{SB}}{1 + 2r\sigma_1^2 c_{11}} \frac{\alpha_1}{a_1},$$

$$t_{2,D}^{SB} = \frac{m_D^{SB}}{1 + 2r\sigma_2^2 c_{22}} \frac{\alpha_2}{a_2}.$$

The larger the variance in an effort dimension, the lower the optimum power of the according incentive. Hence, optimum incentives are weaker in the teaching of soft skills *than they are in hard skills. If $a_i < \alpha_i$, i.e. if no signal is able to represent a student's full capability, the optimum incentive power is the larger, the more actual productivity deviates from the signal.*

The resulting equilibrium values of e_1 and e_2 are

$$e_{1,D}^{SB} = e_1^* \frac{1}{1 + 2r\sigma_1^2 c_{11}},$$
$$e_{2,D}^{SB} = e_2^* \frac{1}{1 + 2r\sigma_2^2 c_{22}},$$

which implies that in the absence of risk (or if the teacher is risk neutral), effort levels are first-best even in the multitasking framework. Note that optimum efforts no longer depend on the signal production parameters a_1, a_2: The distortion of the signal is fully compensated by the appropriate power of the incentive contract. If teachers are risk-averse, the downward distortion in an effort dimension is the larger, the larger the noise in the according signal. In such a situation, the use of an aggregate signal leads to additional inefficient effort contraction.

3.5 Teachers' Own Motivation

So far, we have assumed that teachers are not motivated in the sense that they derive no utility from successful students apart from their monetary remuneration. It is most often the case, however, that teachers choose their profession exactly for the reason of their motivation. Spear, Gould, and Lee (2000) present evidence that a teaching career scores highly for undergraduates on the opportunities given for having creative input, benefiting society, and working with individuals. The most common reasons are job satisfaction and working with children. The reasons rated as least important included working hours, holidays, salaries and security. It seems that prospective teachers are principally attracted to the profession by the rewarding nature of the work involved, as opposed to the pay or conditions on offer.

It is obvious that teacher motivation and morale are of eminent importance in determining students' educational achievement. Studies analyzed by Spear, Gould, and Lee (2000) reveal that teachers believe their own morale to be largely determined by their quality of life within the school, rating factors such as good relations with pupils and helping pupils to achieve as very important. When asked to name those factors that they felt could have a positive effect on the morale of the profession as a whole, teachers' responses largely relate to factors external to the process of teaching itself, focusing on a more positive portrayal of the teaching profession by the media, increased pay and conditions and less pressure. It seems that to improve both the morale of individual teachers and the ethos of the profession as a whole, a range of measures is needed, addressing both experiences integral to the work of teaching, and factors linked to the structural and social context within which that work is carried out.

The main factor found to contribute to the job satisfaction of teachers is working with children. Additional factors included developing warm, personal relationships with pupils, the intellectual challenge of teaching as well as autonomy and independence. In contrast, teachers viewed job dissatisfaction as principally contributed to by work overload, poor pay and perceptions of how teachers are viewed by society.

To experience high job satisfaction, teachers need an intellectual challenge, their autonomy, to feel that they are benefiting society, to enjoy good relations with their colleagues and to spend a sufficient proportion of their time working with children. Enhanced pay, improved status, a less demanding workload and fewer administrative responsibilities should result in lower levels of job dissatisfaction among teachers, but will not necessarily bring about higher levels of job satisfaction.

We operationalize the concept of self-motivation by assuming that every teacher $l \in L$ (L being the set of all teachers) derives an additional private benefit from teaching

$$B_l = \pi^m \sum_i \beta_{l,i} e_i,$$

where subscript i denotes various effort dimensions. Result 10 states the properties of optimum incentive contracts in the case that teachers derive a direct personal benefit from teaching.

Result 10 *If teachers are self-motivated,*
(a) the optimum incentive contract substitutes self-motivation. This means that effort dimensions in which a teacher has a high degree of self-motivation should be rewarded less generously than dimensions with low self-motivation.
(b) the incentive contract simultaneously serves as tool to ensure self-selection of highly motivated teachers into teaching contracts.

Proof. The results correspond to equations (26) and (29)in the appendix. ■

Result 10b follows from the binding participation constraint. If a teacher derives less self-motivation from her work than presumed by the school authority, her participation constraint will not be satisfied. This result is in line with Lazear (1999, 2000), who argues that variable pay is often used as a selection device, rather than as an incentive.

Example 11 *(1 cont.) We illustrate statement (a) of result 10 using the same framework as above. The teacher is now assumed to derive a private benefit* $B = \pi^m (\beta_1 e_1 + \beta_2 e_2)$ *from teaching (or her students' success). Given the power of incentives,* t_1 *and* t_2*, her optimum efforts therefore write as*

$$e_1 = \frac{1}{2}\pi^m \frac{t_1 a_1 + \beta_1}{c_{11}},$$
$$e_2 = \frac{1}{2}\pi^m \frac{t_2 a_2 + \beta_2}{c_{22}}.$$

This is taken into account for the design of the optimum contract which fulfills

$$t_1 a_1 + \beta_1 = \frac{m\alpha_1 + \beta_1}{1 + 2r\sigma_1^2 c_{11}},$$
$$t_2 a_2 + \beta_2 = \frac{m\alpha_2 + \beta_2}{1 + 2r\sigma_2^2 c_{22}}.$$

Hence, as long as $2r\sigma_i^2 c_{ii} > 0$*, the power of the incentive t in a certain dimension decreases in a teacher's own motivation.*

3.6 Educational Production in Teams

We have already addressed the possibility of team-based incentives before. In this section, we explicitly take into account that several teachers are involved in the educational process. For simplicity, we assume that each teacher in a team has control over just one effort dimension.[6] The creation of school-based incentives is usually preferred to individual incentives since these are supposed to be more conductive to cooperation among teachers (cf. Hanushek, 1996, and Craig and Sheu, 1992). In order to encourage cooperation among teachers within schools

[6]This assumption is not critical: If a teacher affects student learning in several dimensions, these can be subsumed under one common aggregate dimension.

and to avoid creating incentives for teachers to sabotage each others' effort, merit pay may be based on the performance of all pupils in a school, with each subject equally weighted, rather than on a teacher-by-teacher basis.[7] We compare the outcomes of two incentive schemes: One in which teachers are paid individually and one in which teachers have a collective contract.

Individual Contracts If teachers are paid individually, they are assumed to behave egoistically and dispense with cooperation. In our model framework, this amounts to letting potential gains of cooperation via the interaction of effort cost c_{ij} lie idle. Hence, every teacher considers her own effort costs and ignores positive spillover effects on her colleagues. However, the advantage of individual merit pay comes from the possibility of providing different effort dimensions and hence different teachers with differently powered incentives.

Collective Contract If teachers are paid collectively, all incentive dimensions must be equally powered, hence diverging from the optimum contract. On the other hand, as a team, teachers cooperate and let others profit from positive spillovers in the educational production process.

The decision whether or not to incentivize teachers individually or in teams depends on the trade-off between effort cooperation and incentive tuning which yields

Result 12 *Individual incentives dominate collective merit pay if spillover effects are small and if the signal distortion is large.*

Proof. The result corresponds to equation (30) in the appendix. ∎

Example 13 *Large potential gains from cooperation among teachers should be realized by paying teachers in teams in order to support teamwork. However, if test measures strongly deviate from actual student performance, teachers are better paid individually in order to profit from the possibility of specifically adjusting incentives in every single dimension.*

4 Conclusion

This paper has studied the properties of optimum incentive contracts in schools with multiple effort dimensions. Our objective has been to identify the critical

[7] This is exactly how a program works which was conducted in western Kenya by the Dutch NGO International Christelijk Steunfonds (cf. for its assessment Glewwe, Ilias, and Kremer, 2003).

characteristics of teachers and school institutions which determine the optimum incentive structures and to identify possible reasons for the absence of incentives contracts in schools. In particular, the model identifies the following distinct forces that shape the optimum contract and determine its applicability in schools:

Distortion of the performance measure The larger the distortion of the performance measure (given by the deviation of the measure from actual performance), the lower the optimum power of incentives.

Noise in the performance measure The noisier the performance measure, the lower the optimum power of incentives. If teachers are risk-averse, high powered incentives demand high risk premia to meet their participation constraint.

Degree of aggregation in the performance measure The higher the performance measure is aggregated, the less specific incentives work and the lower is the optimum power of incentives.

Coordination of efforts Situations in which there are positive cross-effort externalities with educational production favor incentives based on teacher team work.

Teacher attitude towards risk The more risk averse teachers are, the lower is the optimum power of incentives. Again, risk-aversion constitutes the need for a compensation of risk-bearing, which is socially costly.

Teacher motivation With an appropriate incentive structure, preexisting teacher motivation can be benefited from. Potential selection problems do not materialize due to binding participation constraints.

The analysis of teacher motivation and incentives in schools against the background of the empirical evidence tells a cautionary tale about the introduction of incentive schemes based on merit pay. If not well designed, such programs are likely to degrade preexisting motivation and lead to effort substitution into areas which are well measurable but virtually useless with respect to the very goals of education. Since the achievement of educational goals is extremely hard to measure, the optimum power of incentives in education is bound to be lower than in other professions.

5 Appendix

The appendix gives the formal exposition of the models discussed in the body of the paper and proofs to the results stated therein.

The social surplus of education is given by educational production (the numéraire good) minus teacher effort cost; teachers' utility is assumed to be of the form

$$U(x) = -\exp(-xr) \tag{4}$$

where x is the teachers' net benefit (salary minus effort cost plus possibly private benefits from successful students) and r is the Arrow-Pratt measure of absolute risk-aversion.

5.1 Observability of Teacher Efforts

Education is produced according to

$$P(\mathbf{e},m) = \pi^m \sum_{i=1}^{n} \alpha_i e_i = \pi^m \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}' \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix} = \pi^m \mathbf{f}\mathbf{f}'\mathbf{e}$$

where m denotes class size, n is the number of dimensions of teacher effort, α_i is the marginal productivity of input i which is employed at quantity e_i and π is a student's probability of non-disruption in class. On the cost side, various efforts are allowed to interact with each other. The effort cost function writes as

$$C(\mathbf{e}) = \sum_{i=1}^{n} \sum_{j=1}^{n} e_i c_{ij} e_j = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}' \begin{pmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{n1} & \cdots & c_{nn} \end{pmatrix} \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix} = \mathbf{e}'\mathbf{\Gamma}\mathbf{e}$$

where $\mathbf{\Gamma}$ is assumed to be symmetric and positive definite.[8] The entirety of off-diagonal elements of $\mathbf{\Gamma}$ captures the interactions between various effort dimensions. We assume that the different effort dimensions complement each other, i.e. $c_{ij} < 0 \forall i \neq j$.

[8]The assumption of positive definiteness is economically justified by the requirement that effort in any dimension be costly.

If educational production and teacher efforts are observable, the school authority faces the following per-student problem Π^*:

$$\Pi^* : \qquad W^* = \max_{m \in \mathbb{R}_+, \mathbf{e} \in \mathbb{R}_+^n} \left\{ P\left(m, \mathbf{e}\right) - \frac{1}{m} C\left(\mathbf{e}\right) \right\}.$$

Exploiting the first-order conditions[9] yields the optimum values of teacher effort and class size

$$\mathbf{e}^* = \frac{1}{2} m^* \pi^{m^*} \mathbf{\Gamma}^{-1} \mathbf{ff}, \tag{5}$$

$$m^* = -\frac{1}{2 \ln \pi}. \tag{6}$$

Inserting the optimum values into the definition of surplus yields the maximum attainable welfare[10]

$$\begin{aligned} W^* &= \frac{1}{2} m^* \pi^{2m^*} \alpha' \mathbf{\Gamma}^{-1} \mathbf{ff} - \frac{1}{4} m \pi^{2m^*} \mathbf{ff}' \mathbf{\Gamma}^{-1} \mathbf{ff} \\ &= -\frac{1}{8 \ln\left(\pi\right) \epsilon} \mathbf{ff}' \mathbf{\Gamma}^{-1} \mathbf{ff}. \end{aligned} \tag{7}$$

5.2 Output-based Aggregate Signal

In the following, we consider the case that teacher effort is not observable and that educational production is only observable via a possibly distorted and noisy signal $S\left(m, \mathbf{e}\right)$. A teachers's pay T now consists of a base salary b and a signal-depending part $tS\left(m, \mathbf{e}\right)$, where t is the slope of the compensation schedule and can be interpreted as the power of the applied incentive contract:

$$T\left(S\left(m, \mathbf{e}\right)\right) = b + tS\left(m, \mathbf{e}\right).$$

Without direct private benefits form teaching, a teacher's net benefit hence writes as $x = T - C$. We restrict ourselves to linear transfers which are often used in the literature, although – as Holmström and Milgrom (1987) show – the conditions for such a scheme to be optimal are quite stringent. We will also use them in the following since they allow for a simple intuitive interpretation of the results.

[9] Note that $\frac{\partial}{\partial \mathbf{x}} \left(\breve{}' \mathbf{x} \right) = \breve{}$ and $\frac{\partial}{\partial \mathbf{x}} \left(\mathbf{x}' \mathbf{\Lambda} \mathbf{x} \right) = 2 \mathbf{\Lambda} \mathbf{x}$.

[10] The symbol ϵ denotes Euler's number.

The signal $S\,(m,\mathbf{e})$ is produced by a similar technology as actual performance:

$$\begin{aligned} S\,(m,\mathbf{e}) &= \pi^m \sum_{i=1}^{n} s_i\,(e_i) = \pi^m \sum_{i=1}^{n} a_i\,(e_i+\varepsilon_i) \\ &= \pi^m \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}' \left[\begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix} \right] = \pi^m \mathbf{a}'\,(\mathbf{e}+{''})\,. \end{aligned}$$

The vector of errors $'' \equiv (\varepsilon_1,\ldots,\varepsilon_n)'$ is distributed $N\,(0,\mathbf{\Sigma})$; the covariance matrix $\mathbf{\Sigma}$ is symmetric by definition.[11] Note that $S\,(m,\mathbf{e})$ is an aggregate signal in the sense that the various contributions of efforts to production are not disclosed individually. The school authority's design of the incentive contracts must account for the following constraints:

$$\text{PC}: \quad E\,[U\,(b+tS\,(m,\mathbf{e})-C\,(\mathbf{e}))] \geq 0, \tag{8}$$

$$\text{IC}: \quad \mathbf{e} = \arg\max_{\tilde{\mathbf{e}}\in\mathbb{R}_+^n} \{E\,[U\,(b+tS\,(m,\tilde{\mathbf{e}})-C\,(\tilde{\mathbf{e}}))]\}\,. \tag{9}$$

The participation constraint (PC) guarantees that teachers are willing to enter the profession in the first place by assuring them a threshold utility equal to their outside option which is assumed to be equal to zero. The incentive constraint (IC) regards the optimum reaction of teachers to the proposed contract. Following the standard procedure, we assign the role of the mechanism designer to the school authority as principal (cf. eg. Macho-Stadler and Pérez-Castrillo, 1997) whose per-student problem Π^{SB} hence writes as

$$\Pi^{SB}: \qquad W^{SB} = \max_{b,m\in\mathbb{R}_+,t\in\mathbb{R}} \left\{ P\,(m,\mathbf{e}) - \frac{1}{m}\,(b+tS\,(m,\mathbf{e})) \right\}$$
$$\text{subject to (8) and (9).}$$

Then, the fixed part of the teacher's total salary, b, compensates for uncertainty,[12]

$$b = C\,(\mathbf{e}) - t\pi^m\mathbf{a}'\mathbf{e} + \left(rt^2\pi^{2m} \big/ 2 \right) \mathbf{a}'\mathbf{\Sigma}\mathbf{a} \tag{10}$$

[11] The covariance matrix $\mathbf{\Sigma}$ is $\begin{pmatrix} \sigma_1^2 & \cdots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \cdots & \sigma_n^2 \end{pmatrix}$.

[12] Cf. Salanié (1994) and the discussion in footnote 4.

and the per-student problem simplifies to

$$\Pi_A^{SB}: \qquad W_A^{SB} = \max_{m\in\mathbb{R}_+,t\in\mathbb{R}} \left\{ P\left(m,\mathbf{e}\right) - \frac{1}{m}\mathbf{e}'\boldsymbol{\Gamma}\mathbf{e} - \frac{rt^2\pi^{2m}}{2m}\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a} \right\}$$
$$\text{subject to (9).}$$

Since the teacher's utility is monotonic in the received salary minus effort costs, she chooses efforts so as to maximize this difference. Using (10) in (9), she solves the problem

$$\mathbf{e} \in \arg\max_{\tilde{\mathbf{e}}\in\mathbb{R}_+^n} \left\{ t\pi^m\mathbf{a}'\tilde{\mathbf{e}} - \tilde{\mathbf{e}}'\boldsymbol{\Gamma}\tilde{\mathbf{e}} \right\}. \tag{11}$$

The optimum effort vector can be determined by the first-order condition.[13] Solving for $\mathbf{e}$ yields

$$\mathbf{e} = \frac{t}{2}\pi^m\boldsymbol{\Gamma}^{-1}\mathbf{a}.$$

The reduced form of the school authority's decision problem $\Pi_A'^{SB}$ hence writes as

$$\Pi_A'^{SB}: \qquad W_A^{SB} = \max_{m\in\mathbb{R}_+,t\in\mathbb{R}} \left\{ \frac{t}{2}\pi^{2m}\mathbf{ff}'\boldsymbol{\Gamma}^{-1}\mathbf{a} - \frac{t^2}{4m}\pi^{2m}\mathbf{a}'\boldsymbol{\Gamma}^{-1}\mathbf{a} - \frac{rt^2\pi^{2m}}{2m}\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a} \right\}.$$

The optimum second-best values of the power of incentives and class size are derived directly from the first-order conditions and given by

$$\begin{aligned} t_A^{SB} &= m_A^{SB}\frac{\mathbf{ff}'\boldsymbol{\Gamma}^{-1}\mathbf{a}}{\mathbf{a}'\boldsymbol{\Gamma}^{-1}\mathbf{a}+2r\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}}, \\ m_A^{SB} &= -\frac{1}{2\ln\pi}. \end{aligned} \tag{12}$$

For a geometrical interpretation set $\boldsymbol{\Gamma} = c\mathbf{I}$, and $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}$ where $\mathbf{I}$ is the identity matrix, such costs for different tasks are equal and there is no interaction between the various effort dimensions. Then, rewriting the numerator in (12)

$$t_A^{SB} = m_A^{SB}\frac{c\left|\mathbf{ff}\right|\left|\mathbf{a}\right|\cos\theta}{\mathbf{a}'\boldsymbol{\Gamma}^{-1}\mathbf{a}+2r\mathbf{a}'\sigma^2\mathbf{I}\mathbf{a}},$$

[13] To justify the use of the first-order approach in this setting, it suffices to show that 11 has a unique maximizer which satisfies the first-order condition. This is the case as $\mathbf{a}'\tilde{\mathbf{e}}$ is linear in $\mathbf{e}$ and $\tilde{\mathbf{e}}'\boldsymbol{\Gamma}\tilde{\mathbf{e}}$ is strictly convex in $\mathbf{e}$ due to the positive definiteness of $\boldsymbol{\Gamma}$. We are hence maximizing a strictly concave function on a convex set. Rogerson (1985) describes sufficient conditions for the first-order approach to be valid in general.

such that, when $\cos\theta$ is positive,

$$
\begin{array}{ll}
\frac{dt_A^{SB}}{d\cos(\theta)} < 0: & \text{the optimum power of incentives decreases in the} \\
 & \text{distortion of the performance measure,} \\
\frac{dt_A^{SB}}{d\sigma^2} < 0: & \text{the optimum power of incentives decreases in the} \\
 & \text{noise of the performance measure,} \\
\frac{dt_A^{SB}}{dr} < 0: & \text{the optimum power of incentives decreases in the} \\
 & \text{teacher's risk aversion,} \\
\frac{dt_A^{SB}}{d\pi} > 0: & \text{the optimum power of incentives increases in} \\
 & \text{the probability that a student behaves well.}
\end{array} \tag{13}
$$

With an output-based performance measure, the optimum second-best effort vector is

$$\mathbf{e}_A^{SB} = \frac{1}{2} m_A^{SB} \pi^{m_A^{SB}} \mathbf{\Gamma}^{-1} \mathbf{a} \frac{\mathbf{ff}'\mathbf{\Gamma}^{-1}\mathbf{a}}{\mathbf{a}'\mathbf{\Gamma}^{-1}\mathbf{a}+2r\mathbf{a}'\mathbf{\Sigma}\mathbf{a}}.$$

The social surplus is given by

$$W_A^{SB} = -\frac{1}{8\ln(\pi)\epsilon} \frac{\left(\mathbf{a}'\mathbf{\Gamma}^{-1}\mathbf{ff}\right)^2}{\mathbf{a}'\mathbf{\Gamma}^{-1}\mathbf{a}+2r\mathbf{a}'\mathbf{\Sigma}\mathbf{a}}. \tag{14}$$

Note that with $\mathbf{ff} = \mathbf{a}$, $r = 0$, and $\sigma_i^2 = 0\ \forall i$, $W_A^{SB} = W^*$. Moreover, with $\mathbf{\Gamma} = c\mathbf{I}$, and $\mathbf{\Sigma} = \sigma^2\mathbf{I}$,

$$\frac{dW_A^{SB}}{d\theta} < 0, \qquad \frac{dW_A^{SB}}{d\sigma^2} < 0, \qquad \frac{dW_A^{SB}}{dr} < 0.$$

Hence, asymmetric information reduces welfare due to the distortion and the noise of the performance measure, but also due to the teacher's risk bearing.

5.3 Input-Based Aggregate Signal

In this section, we consider the case that the signal is independent of the overall behavior of students in class, i.e. the resulting signal is input-based, $S(\mathbf{e}) = \mathbf{a}'(\mathbf{e} + ")$. Again, the fixed part of the teacher's total salary, b, compensates for uncertainty, $b = C(\mathbf{e}) - t\mathbf{a}'\mathbf{e} + \left(rt^2/2\right)\mathbf{a}'\mathbf{\Sigma}\mathbf{a}$. The per-student problem is then

$$\Pi^{SB}: \qquad W^{SB} = \max_{m\in\mathbb{R}_+, t\in\mathbb{R}} \left\{ P(m,\mathbf{e}) - \frac{1}{m}\mathbf{e}'\mathbf{\Gamma}\mathbf{e} - \frac{rt^2}{2m}\mathbf{a}'\mathbf{\Sigma}\mathbf{a} \right\}$$

subject to (9).

The teacher chooses efforts so as to solve the problem

$$\mathbf{e} \in \arg \max_{\tilde{\mathbf{e}} \in \mathbb{R}^n_+} \left\{ t\mathbf{a}'\tilde{\mathbf{e}} - \tilde{\mathbf{e}}'\mathbf{\Gamma}\tilde{\mathbf{e}} \right\}.$$

The maximizer can be determined by the first-order conditions (cf. footnote 13). Solving for $\mathbf{e}$ yields

$$\mathbf{e} = \frac{t}{2}\mathbf{\Gamma}^{-1}\mathbf{a}.$$

The reduced form of the school authority's decision problem hence writes as

$$\Pi'^{SB}: \qquad W^{SB} = \max_{m \in \mathbb{R}_+, t \in \mathbb{R}} \left\{ \frac{t}{2}\pi^m \mathbf{f}\mathbf{f}'\mathbf{\Gamma}^{-1}\mathbf{a} - \frac{t^2}{4m}\mathbf{a}'\mathbf{\Gamma}^{-1}\mathbf{a} - \frac{rt^2}{2m}\mathbf{a}'\mathbf{\Sigma}\mathbf{a} \right\}.$$

From the first-order conditions we get again

$$\begin{aligned} t^{SB} &= m^{SB}\pi^{m^{SB}} \frac{\mathbf{f}\mathbf{f}'\mathbf{\Gamma}^{-1}\mathbf{a}}{\mathbf{a}'\mathbf{\Gamma}^{-1}\mathbf{a} + 2r\mathbf{a}'\mathbf{\Sigma}\mathbf{a}}, \\ m^{SB} &= -\frac{1}{2\ln\pi}, \end{aligned}$$

with the optimum effort vector being

$$\mathbf{e}^{SB} = \frac{1}{2}m^{SB}\pi^{m^{SB}}\mathbf{\Gamma}^{-1}\mathbf{a}\frac{\mathbf{f}\mathbf{f}'\mathbf{\Gamma}^{-1}\mathbf{a}}{\mathbf{a}'\mathbf{\Gamma}^{-1}\mathbf{a} + 2r\mathbf{a}'\mathbf{\Sigma}\mathbf{a}}$$

and the social surplus is

$$W^{SB} = -\frac{1}{8\ln(\pi)\,\epsilon}\frac{\left(\mathbf{a}'\mathbf{\Gamma}^{-1}\mathbf{f}\mathbf{f}\right)^2}{\mathbf{a}'\mathbf{\Gamma}^{-1}\mathbf{a} + 2r\mathbf{a}'\mathbf{\Sigma}\mathbf{a}} = W_A^{SB}. \tag{15}$$

As an alternative, consider the possibility that the input-based aggregate signal is observable without noise, $S(\mathbf{e}) = \mathbf{a}'\mathbf{e}$, but at a cost $\mu C(\mathbf{e})$. Again, the fixed part of a teacher's total wage, b, assures that the participation constraint is met, $b = C(\mathbf{e}) - t\mathbf{a}'\mathbf{e}$. The per-student problem is then

$$\Pi_{NL}^{SB}: \qquad W_{NL}^{SB} = \max_{m \in \mathbb{R}_+, t \in \mathbb{R}} \left\{ P(\mathbf{e}, m) - \frac{1}{m}(1+\mu)\,\mathbf{e}'\mathbf{\Gamma}\mathbf{e} \right\}$$

subject to (9).

The teacher chooses efforts so as to solve the problem

$$\mathbf{e} \in \arg \max_{\tilde{\mathbf{e}} \in \mathbb{R}^n_+} \left\{ t\mathbf{a}'\tilde{\mathbf{e}} - \tilde{\mathbf{e}}'\mathbf{\Gamma}\tilde{\mathbf{e}} \right\}.$$

The maximizer can be determined by the first-order conditions (cf. footnote 13). Solving for $\mathbf{e}$ yields

$$\mathbf{e} = \frac{t}{2}\mathbf{\Gamma}^{-1}\mathbf{a}.$$

The reduced form of the school authority's decision problem hence writes as

$$\Pi_{NL}^{\prime SB}: \qquad W_{NL}^{SB} = \max_{m\in\mathbb{R}_+, t\in\mathbb{R}} \left\{ \frac{t}{2}\pi^m \mathbf{ff}'\mathbf{\Gamma}^{-1}\mathbf{a} - (1+\mu)\frac{t^2}{4m}\mathbf{a}'\mathbf{\Gamma}^{-1}\mathbf{a} \right\}.$$

From the first-order conditions we get again

$$t_{NL}^{SB} = \frac{m_{NL}^{SB}\pi^{m_{NL}^{SB}}}{1+\mu}\frac{\mathbf{ff}'\mathbf{\Gamma}^{-1}\mathbf{a}}{\mathbf{a}'\mathbf{\Gamma}^{-1}\mathbf{a}},$$
$$m_{NL}^{SB} = -\frac{1}{2\ln\pi}$$

and the social surplus is

$$W_{NL}^{SB} = -\frac{1}{8\ln(\pi)\,\epsilon}\frac{\left(\mathbf{a}'\mathbf{\Gamma}^{-1}\mathbf{ff}\right)^2}{(1+\mu)\,\mathbf{a}'\mathbf{\Gamma}^{-1}\mathbf{a}}. \tag{16}$$

Hence, comparing (15) to (16), the use of the noiseless signal is preferred iff

$$\mu < 2r\frac{\mathbf{a}'\mathbf{\Sigma}\mathbf{a}}{\mathbf{a}'\mathbf{\Gamma}^{-1}\mathbf{a}}. \tag{17}$$

5.4 Disaggregate Signal

In this section we analyze the possibility of rewarding different dimensions of effort differently by a vector $\mathbf{t}$ of differentiated incentive powers, $\mathbf{t} \equiv (t_1, \ldots, t_n)'$. This requires that individual inputs be separately measurable. Hence, we have to rely on a disaggregate signal of the form $\mathbf{s}(m, \mathbf{e}) \equiv (s_1(m, e_1), \ldots, s_n(m, e_n))' = \pi^m(\mathbf{a} \bullet (\mathbf{e} + ''))$.[14] The total transfer to teachers is thus

$$T(\mathbf{s}(m,\mathbf{e})) = b + \mathbf{t}'\mathbf{s}(m,\mathbf{e}).$$

Since the input-based and output-based performance measures are equivalent, we analyze the more intuitive output-based one in the following. The participation and incentive constraints write as

$$\text{PC}: \qquad E\left[U\left(b + \mathbf{t}'\mathbf{s}(m,\mathbf{e}) - C(\mathbf{e})\right)\right] \geq 0, \tag{18}$$
$$\text{IC}: \qquad \mathbf{e} \in \arg\max_{\tilde{\mathbf{e}}\in\mathbb{R}_+^n} \left\{E\left[U\left(b + \mathbf{t}'\mathbf{s}(m,\tilde{\mathbf{e}}) - C(\tilde{\mathbf{e}})\right)\right]\right\} \tag{19}$$

[14]The operator $\bullet$ denotes entrywise multiplication of matrices (Hadamard product).

respectively. The welfare maximization problem Π_D^{SB} becomes in analogy to above

$$\Pi_D^{SB}: \qquad W^{SB} = \max_{b\in\mathbb{R}_+,\mathbf{t}\in\mathbb{R}^n} \left\{ P(m,\mathbf{e}) - \frac{1}{m}\left(b + \mathbf{t}'\mathbf{s}(m,\mathbf{e})\right)\right\}$$
$$\text{subject to (18) and (19).}$$

Again, the fixed part of the teacher's salary, b, compensates for uncertainty, $b = C(\mathbf{e}) - \pi^m (\mathbf{t}\bullet\mathbf{a})'\mathbf{e} + \left(r\pi^{2m}/2\right)(\mathbf{t}\bullet\mathbf{a})'\boldsymbol{\Sigma}(\mathbf{t}\bullet\mathbf{a})$, such that the per-student problem becomes

$$\Pi_D^{SB}: \qquad W_D^{SB} = \max_{\mathbf{t}\in\mathbb{R}^n} \left\{ P(m,\mathbf{e}) - \frac{1}{m}\mathbf{e}'\boldsymbol{\Gamma}\mathbf{e} - \frac{r\pi^{2m}}{2m}(\mathbf{t}\bullet\mathbf{a})'\boldsymbol{\Sigma}(\mathbf{t}\bullet\mathbf{a})\right\}$$
$$\text{subject to (19).}$$

The teacher chooses efforts so as to find

$$\mathbf{e} \in \arg\max_{\tilde{\mathbf{e}}\in\mathbb{R}_+^n} \left\{ \pi^m (\mathbf{t}\bullet\mathbf{a})'\tilde{\mathbf{e}} - \tilde{\mathbf{e}}'\boldsymbol{\Gamma}\tilde{\mathbf{e}} \right\}.$$

The maximizing effort vector $\mathbf{e}$ can be determined using the first-order conditions (cf. footnote 13). Solving for $\mathbf{e}$ yields

$$\mathbf{e} = \frac{1}{2}\pi^m \boldsymbol{\Gamma}^{-1}(\mathbf{t}\bullet\mathbf{a}).$$

Hence, the reduced form problem writes as

$$\Pi_D'^{SB}: \qquad W_D^{SB} = \max_{\mathbf{t}\in\mathbb{R}^n} \left\{ \frac{1}{2}\pi^{2m}\mathbf{ff}'\boldsymbol{\Gamma}^{-1}(\mathbf{t}\bullet\mathbf{a}) - \frac{1}{4m}\pi^{2m}(\mathbf{t}\bullet\mathbf{a})'\boldsymbol{\Gamma}^{-1}(\mathbf{t}\bullet\mathbf{a}) - \frac{r}{2m}\pi^{2m}(\mathbf{t}\bullet\mathbf{a})'\boldsymbol{\Sigma}(\mathbf{t}\bullet\mathbf{a}) \right\}.$$

From the first-order conditions we get

$$\mathbf{t}_D^{SB}\bullet\mathbf{a} = m_D^{SB}\left(\mathbf{I} + 2r\boldsymbol{\Gamma}\boldsymbol{\Sigma}\right)^{-1}\mathbf{ff}, \tag{20}$$
$$m_D^{SB} = -\frac{1}{2\ln\pi}. \tag{21}$$

with the optimum effort vector being

$$\mathbf{e}_D^{SB} = \frac{1}{2}m_D^{SB}\pi^{m_D^{SB}}\left(\boldsymbol{\Gamma} + 2r\boldsymbol{\Gamma}\boldsymbol{\Sigma}\boldsymbol{\Gamma}\right)^{-1}\mathbf{ff}$$

and overall welfare is

$$W_D^{SB} = -\frac{1}{8\ln(\pi)\,\epsilon}\mathbf{ff}'\mathbf{X}\mathbf{ff} \tag{22}$$

with

$$\mathbf{X} = \left[2\left(\mathbf{\Gamma}+2r\mathbf{\Gamma\Sigma\Gamma}\right)^{-1} - \left(\mathbf{\Gamma}+4r\mathbf{\Gamma\Sigma\Gamma}+4r^2\mathbf{\Gamma\Sigma\Gamma\Sigma\Gamma}\right)^{-1} - 2r\left(\mathbf{\Gamma}^{-1}+4r\mathbf{\Gamma}+4r\mathbf{\Gamma\Sigma\Gamma}\right)^{-1}\right]. \tag{23}$$

5.5 Motivated Teachers

We model the case of motivated teachers by assuming that their net benefit x in (4) is $T(\mathbf{s}(m,\mathbf{e})) - C(\mathbf{e}) + B(m,\mathbf{e})$ with $B(m,\mathbf{e}) = \pi^m \mathbf{fi}'\mathbf{e}$ denoting the teachers' private benefit from successful students. We further assume that a disaggregate signal is available as in the previous section. Hence, the participation and incentive constraints write as

$$\text{PC}: \quad E\left[U\left(b+\mathbf{t}'\mathbf{s}(m,\mathbf{e}) - C(\mathbf{e}) + \pi^m\mathbf{fi}'\mathbf{e}\right)\right] \geq 0, \tag{24}$$

$$\text{IC}: \quad \mathbf{e} \in \arg\max_{\tilde{\mathbf{e}}\in\mathbb{R}^n_+}\left\{E\left[U\left(b+\mathbf{t}'\mathbf{s}(m,\tilde{\mathbf{e}}) - C(\tilde{\mathbf{e}}) + \pi^m\mathbf{fi}'\mathbf{e}\right)\right]\right\} \tag{25}$$

respectively. The welfare maximization problem Π_D^{SB} becomes in analogy to above

$$\Pi_M^{SB}: \quad W_M^{SB} = \max_{b\in\mathbb{R},\mathbf{t}\in\mathbb{R}^n}\left\{P(m,\mathbf{e}) - \frac{1}{m}\left(b+\mathbf{t}'\mathbf{s}(m,\mathbf{e})\right)\right\}$$
$$\text{subject to (24) and (25).}$$

The fixed part of the teacher's salary, b, compensates for uncertainty,

$$b = C(\mathbf{e}) - \pi^m\left(\mathbf{t}\bullet\mathbf{a}+\mathbf{fi}\right)'\mathbf{e} + \frac{r\pi^{2m}}{2}\left(\mathbf{t}\bullet\mathbf{a}+\mathbf{fi}\right)'\mathbf{\Sigma}\left(\mathbf{t}\bullet\mathbf{a}+\mathbf{fi}\right),$$

such that the per-student problem becomes

$$\Pi_M^{SB}: \quad W_M^{SB} = \max_{\mathbf{t}\in\mathbb{R}^n}\left\{P(m,\mathbf{e}) + \frac{1}{m}B(m,\mathbf{e}) - \frac{1}{m}\mathbf{e}'\mathbf{\Gamma}\mathbf{e} - \frac{r\pi^{2m}}{2m}\left(\mathbf{t}\bullet\mathbf{a}+\mathbf{fi}\right)'\mathbf{\Sigma}\left(\mathbf{t}\bullet\mathbf{a}+\mathbf{fi}\right)\right\}$$
$$\text{subject to (25).}$$

The teacher chooses efforts so as to find

$$\mathbf{e} \in \arg\max_{\tilde{\mathbf{e}}\in\mathbb{R}^n_+}\left\{\pi^m\left(\mathbf{t}\bullet\mathbf{a}+\mathbf{fi}\right)'\tilde{\mathbf{e}} - \tilde{\mathbf{e}}'\mathbf{\Gamma}\tilde{\mathbf{e}}\right\}.$$

The maximizing effort vector $\mathbf{e}$ can be determined using the first-order conditions (cf. footnote 13). Solving for $\mathbf{e}$ yields

$$\mathbf{e} = \frac{1}{2}\pi^m\mathbf{\Gamma}^{-1}\left(\mathbf{t}\bullet\mathbf{a}+\mathbf{fi}\right).$$

Total welfare consists of the productive value of a student's skills plus her teacher's private value minus the risk premia attached to these two benefit elements. Hence, the reduced form problem writes

$$\Pi_M'^{SB}: \qquad W_M^{SB} = \max_{\mathbf{t}\in\mathbb{R}^n} \left\{ \frac{1}{2}\pi^{2m}\mathbf{ff}'\boldsymbol{\Gamma}^{-1}\left(\mathbf{t}\bullet\mathbf{a}+\mathbf{fi}\right) + \frac{1}{2m}\pi^{2m}\mathbf{fi}'\boldsymbol{\Gamma}^{-1}\left(\mathbf{t}\bullet\mathbf{a}+\mathbf{fi}\right) - \frac{1}{4m}\pi^{2m}\left(\mathbf{t}\bullet\mathbf{a}+\mathbf{fi}\right)'\boldsymbol{\Gamma}^{-1}\left(\mathbf{t}\bullet\mathbf{a}+\mathbf{fi}\right) - \frac{r}{2m}\pi^{2m}\left(\mathbf{t}\bullet\mathbf{a}+\mathbf{fi}\right)'\boldsymbol{\Sigma}\left(\mathbf{t}\bullet\mathbf{a}+\mathbf{fi}\right) \right\}.$$

From the first-order conditions we get

$$\mathbf{t}_M^{SB}\bullet\mathbf{a}+\mathbf{fi} = m_M^{SB}\left(\mathbf{I}+2r\boldsymbol{\Gamma}\boldsymbol{\Sigma}\right)^{-1}\left(\mathbf{ff}+\frac{1}{m}\mathbf{fi}\right), \tag{26}$$

$$m_M^{SB} = -\frac{1}{2\ln\pi} \tag{27}$$

with the optimum effort vector being

$$\mathbf{e}_M^{SB} = \frac{1}{2}m_M^{SB}\pi^{m_M^{SB}}\left(\boldsymbol{\Gamma}+2r\boldsymbol{\Gamma}\boldsymbol{\Sigma}\boldsymbol{\Gamma}\right)^{-1}\left(\mathbf{ff}+\frac{1}{m}\mathbf{fi}\right)$$

and overall welfare is

$$W_M^{SB} = -\frac{1}{8\ln(\pi)\epsilon}\left(\mathbf{ff}+\frac{1}{m}\mathbf{fi}\right)'\mathbf{X}\left(\mathbf{ff}+\frac{1}{m}\mathbf{fi}\right), \tag{28}$$

where $\mathbf{X}$ is given in (23). Recall (24) and note that since

$$\frac{d}{d\beta_i}E\left[U\left(b+\mathbf{t}'\mathbf{s}\left(m,\mathbf{e}\right)-C\left(\mathbf{e}\right)+\pi^m\mathbf{fi}'\mathbf{e}\right)\right] < 0, \tag{29}$$

no teacher with a lower motivation in at least one effort dimension will accept the proposed contract.

5.6 Team-Based Incentives

Up to now, we have studied differentiated efforts by one single teacher. In the following, we analyze different effort dimensions exerted by different teachers. By assumption, we allow for the two possibilities that (1) teachers are paid on an individual basis and hence may compete against each other and (2) they are awarded as a team and also behave as such.[15]

When teachers behave as a team, the problem is equivalent with a single teacher being assessed via one single output based performance measure; also the results for optimum transfer and class size as well as welfare remain unchanged:

$$t_{TA}^{SB} = m_{TA}^{SB} \frac{\mathbf{ff}'\mathbf{\Gamma}^{-1}\mathbf{a}}{\mathbf{a}'\mathbf{C}^{-1}\mathbf{a}+2r\mathbf{a}'\mathbf{\Sigma}\mathbf{a}},$$
$$m_{TA}^{SB} = -\frac{1}{2\ln\pi},$$
$$W_{TA}^{SB} = -\frac{1}{8\ln(\pi)\epsilon} \frac{\left(\mathbf{a}'\mathbf{\Gamma}^{-1}\mathbf{ff}\right)^2}{\mathbf{a}'\mathbf{\Gamma}^{-1}\mathbf{a}+2r\mathbf{a}'\mathbf{\Sigma}\mathbf{a}}.$$

If teachers are treated individually, they can be incentivized with different power and their rewards are independent from each other but they don't let their colleagues profit from their own effort, such that they take into account only their own effort costs and neglect potential positive externalities. Hence, the problem writes as above with disaggregate signals and the solution is

$$\mathbf{t}_{TD}^{SB} \bullet \mathbf{a} = m_D^{SB} \left(\mathbf{I} + 2r\overline{\mathbf{\Gamma}\mathbf{\Sigma}}\right)^{-1} \mathbf{ff},$$
$$m_{TD}^{SB} = -\frac{1}{2\ln\pi}$$

with $\overline{\mathbf{\Gamma}} = diag(\mathbf{\Gamma})$, $\overline{\mathbf{\Sigma}} = diag(\mathbf{\Sigma})$ which takes into account the strict independence of efforts. Total surplus is

$$W_{TD}^{SB} = -\frac{1}{8\ln(\pi)\epsilon}\mathbf{ff}'\overline{\mathbf{X}}\mathbf{ff}$$

with $\overline{\mathbf{X}} = \left[2\left(\overline{\mathbf{\Gamma}}+2r\overline{\mathbf{\Gamma}\mathbf{\Sigma}\mathbf{\Gamma}}\right)^{-1} - \left(\overline{\mathbf{\Gamma}}+4r\overline{\mathbf{\Gamma}\mathbf{\Sigma}\mathbf{\Gamma}}+4r^2\overline{\mathbf{\Gamma}\mathbf{\Sigma}\mathbf{\Gamma}\mathbf{\Sigma}\mathbf{\Gamma}}\right)^{-1} - 2r\left(\overline{\mathbf{\Gamma}}^{-1} + 4r\overline{\mathbf{\Gamma}}+4r\overline{\mathbf{\Gamma}\mathbf{\Sigma}\mathbf{\Gamma}}\right)^{-1}\right]$.

Comparing the results in the two regimes, we consider the case of risk-neutral teachers, such that the resulting welfare measures are given by

$$W_{TA}^{SB} = -\frac{1}{8\ln(\pi)\epsilon} \frac{\left(\mathbf{a}'\mathbf{\Gamma}^{-1}\mathbf{ff}\right)^2}{\mathbf{a}'\mathbf{\Gamma}^{-1}\mathbf{a}},$$
$$W_{TD}^{SB} = -\frac{1}{8\ln(\pi)\epsilon}\mathbf{ff}'\overline{\mathbf{\Gamma}}^{-1}\mathbf{ff},$$

[15] In the second case, we assume social control among teachers, such that there is no freeriding on other teachers' efforts. Of course, this is an extreme assumption, but it serves the very purpose of the argument comparing two diametrically opposed cases.

respectively. The condition for individual merit pay being superior to group incentives is

$$
\begin{aligned}
& W_{TA}^{SB} < W_{TD}^{SB} \\
\Leftrightarrow\ & \frac{\left(\mathbf{a}'\boldsymbol{\Gamma}^{-1}\mathbf{ff}\right)^2}{\mathbf{a}'\boldsymbol{\Gamma}^{-1}\mathbf{aff}'\overline{\boldsymbol{\Gamma}}^{-1}\mathbf{ff}} < 1. \qquad (30)
\end{aligned}
$$

References

BAKER, G. P. (2002): "Distortion and Risk in Optimal Incentive Contracts," *Journal of Human Resources*, 37(4), 728–751.

EBERTS, R., K. HOLLENBECK, AND J. STONE (2002): "Teacher Performance Incentives And Student Outcomes," *Journal of Human Resources*, 37(4), 913–927.

FEHR, E., AND K. M. SCHMIDT (2004): "Fairness and Incentives in a Multi-Tasking Principal-Agent Model," *Scandinavian Journal of Economics*, 106(3), 453–474.

GLEWWE, P., N. ILIAS, AND M. KREMER (2003): "Teacher Incentives," NBER *Working Paper 9671*.

HANNAWAY, J. (1996): "Management Decentralization and Performance-Based Incentives: Theoretical Considerations for Schools," in *Improving America's Schools: The Role of Incentives*, ed. by E. A. Hanushek, and D. W. Jorgenson, pp. 97–109, Washington, D.C. National Academy Press.

HANUSHEK, E. A. (1994): *Making Schools Work: Improving Performance and Controlling Costs*. Brookings Institution Press, Washington, D.C.

——— (1996): "Outcomes, Cost, and Incentives in Schools," in *Improving America's Schools: The Role of Incentives*, ed. by E. A. Hanushek, and D. W. Jorgenson, pp. 29–52, Washington, D.C. National Academy Press.

HANUSHEK, E. A., AND M. E. RAYMOND (2004): "Does School Accountability Lead to Improved Student Performance," NBER *Working Paper 10591*.

HOLMSTRÖM, B., AND P. MILGROM (1987): "Aggregation and Linearity in the Provision of Intertemporal Incentives," *Econometrica*, 55(2), 303–328.

——— (1991): "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics and Organization*, 7(Special Issue), 24–52.

JAAG, C. (2005): "Hidden Teacher Effort in Educational Production: Monitoring vs. Merit Pay," *University of St. Gallen* IFF *Working Paper*.

JACOB, B. A. (2002): "Accountability, Incentives and Behavior: The Impact of High-Stakes Testing in the Chicago Public Schools," NBER *Working Paper 8968*.

JÜRGES, H., W. F. RICHTER, AND K. SCHNEIDER (2004): "Teacher Quality and Incentives: Theoretical and Empirical Effects of Standards on Teacher Quality," CES*ifo Working Paper 1296*.

KANE, T. J., AND D. O. STAIGER (2001): "Improving School Accountability Measures," NBER *Working Paper 8156*.

KORETZ, D. M. (2002): "Limitations in the Use of Achievement Tests as Measures of Educator's Productivity," *Journal of Human Resources*, 37(4), 752–777.

LADD, H. F. (1999): "The Dallas School Accountability and Incentive Program: An Evaluation of its Impacts on Student Outcomes," *Economics of Education Review*, 18, 1–16.

LAVY, V. (2002): "Evaluating the Effect of Teachers' Group Performance Incentives on Pupil Achievement," *Journal of Political Economy*, 110(6), 1286–1317.

——— (2003): "Paying for Performance: The Effect of Teachers' Financial Incentives on Students' Scholastic Outcomes," CEPR *Discussion Paper 3862*.

——— (2004): "Performance Pay and Teachers' Effort, Productivity and Grade Ethics," NBER *Working Paper 10622*.

LAZEAR, E. P. (1999): "Output-Based Pay: Incentives of Sorting?," NBER *Working Paper 7419*.

——— (2000): "The Power of Incentives," *American Economic Review Papers and Proceedings*, 90(2), 410–414.

MACHO-STADLER, I., AND D. PÉREZ-CASTRILLO (1997): *An Introduction to the Economics of Information*. Oxford University Press, Oxford.

PRENDERGAST, C. (2000): "What Trade-Off of Risk and Incentives?," *American Economic Review Papers and Proceedings*, 90(2), 421–425.

RICHARDS, C. E., AND T. M. SHEU (1992): "The South Carolina School Incentive Reward Program: A Policy Analysis," *Economics of Education Review*, 11(1), 71–86.

ROGERSON, W. P. (1985): "The First-Order Approach to Principal-Agent Problems," *Econometrica*, 53(6), 1357–1368.

SALANIÉ, B. (1994): *Théorie des contracts*. Economica, Paris.

SPEAR, M., K. GOULD, AND B. LEE (2000): *Who Would Be a Teacher? A Review of Factors Motivating and Demotivating Prospective and Practising Teachers*. NFER, Slough.