ZEW
Zentrum für Europäische
Wirtschaftsforschung GmbH

Centre for European
Economic Research

Universität
Konstanz

Rechts-, Wirtschafts- und
Verwaltungswissenschaftliche
Sektion
Fachbereich
Wirtschaftswissenschaften

Diskussionspapiere der DFG-
Forschergruppe (Nr.: 3468269275):

Heterogene Arbeit: Positive und Normative
Aspekte der Qualifikationsstruktur der Arbeit

Gerald Eisenkopf

Deterring gaming with imperfect evaluation methods

Januar 2006

Diskussionspapier Nr. 06/04
http://www.wiwi.uni-konstanz.de/forschergruppewiwi/

Diskussionspapier der Forschergruppe (Nr.: 3468269275) "Heterogene Arbeit: Positive und Normative Aspekte der Qualifikationsstruktur der Arbeit"

Nr. 06/04, Januar 2006

Deterring gaming with imperfect evaluation methods

Gerald Eisenkopf

Universität Konstanz
Fachbereich Wirtschaftswissenschaften
Fach D144
78457 Konstanz
Germany

mail: gerald.eisenkopf@uni-konstanz.de
phone: +49-7531-88-3455
fax +49-7531-88-4456

Zusammenfassung:

This paper introduces and discusses an idea which minimizes gaming or manipulation activities, if payments are linked to results from manipulative methods. The idea is to add nonmanipulable information to manipulable information to improve the evaluation of a given output. A score declining in increasing evaluation quality indicates gaming and allows to estimate the true result. A simple linear incentive scheme is introduced in which a high evaluation score is rewarded. The introduced mechanism dominates any single evaluation method. However, limited liability restricts its applicability. If agents are risk-averse, the principal should let each agent decide, which evaluation method he prefers.

JEL Klassifikation : D80, M40, M52
Schlüsselwörter : Manipulation, Gaming, Evaluation, Risk Aversion, Agency Theory, Research Management
Download/Reference : http://www.wiwi.uni-konstanz.de/forschergruppewiwi/

Deterring gaming with imperfect evaluation methods

Gerald Eisenkopf
University of Konstanz*†

January 20, 2006

Abstract

This paper introduces and discusses an idea which minimizes gaming or manipulation activities, if payments are linked to results from manipulative methods. The idea is to add nonmanipulable information to manipulable information to improve the evaluation of a given output. A score declining in increasing evaluation quality indicates gaming and allows to estimate the true result. A simple linear incentive scheme is introduced in which a high evaluation score is rewarded. The introduced mechanism dominates any single evaluation method. However, limited liability restricts its applicability. If agents are risk-averse, the principal should let each agent decide, which evaluation method he prefers.

JEL classifications: D80, M40. M52

Keywords: Manipulation, Gaming, Evaluation, Risk Aversion, Agency Theory, Research Management

*Corresponding Address: University of Konstanz, Box D 144, 78457 Konstanz, Germany, E-Mail: Gerald.Eisenkopf@uni-konstanz.de

†The author thanks Oliver Fabel and seminar participants in Konstanz and Tübingen for helpful discussions. Financial support of the Deutsche Forschungsgemeinschaft (DFG) through the research group “Heterogeneous Labor: Positive and Normative Aspects of the Skill Structure of Labor” is gratefully acknowledged. Comments are very welcome. The usual disclaimer applies.

1 Introduction

The performance of organizations and employees is difficult to measure in many respects. The output is not immediately observable and it is difficult to quantify. The specific impact of an agent's effort is hard to identify. Hence, instruments of performance measurement use observable and inherently superficial characteristics. From these characteristics an evaluator can conclude on the real output (e.g. in academics, publication records often serve as a proxy for research quality). The application of superficial characteristics creates problems if principals link payments to performance. Agents have an interest in creating the observable characteristics. Such measures may incidentally improve desired output but they channel inputs in wasteful, window-dressing activities. Any principal get what he pays for, which is not necessarily what he wants (Kerr, 1975).

To address this problem, Baker, Gibbons and Murphy (1994) propose subjective performance measures while Murphy and Oyer (2003) call for discretion in incentive contracts. However, these instruments may not be feasible in many organizations, e.g. in the public sector, where evaluation and payments should be as transparent as possible to restrict corruption. Courty and Marschke (CM; 2003) recommend to take the dynamic characteristics of gaming into account, as agents learn the mechanisms of the incentive contract much more precisely than the principal can anticipate. Hence, gaming increases over time and requires a change in the performance measurement. Their paper also includes a very helpful review of the literature.

In this paper, I propose another mechanism to minimize the wasteful activities if payment is linked to a more or less superficial evaluation method. Like in CM agents can game the results but in this paper gaming affects all evaluation methods. Then the idea is to increase evaluation quality by considering more information for the performance measurement. Hence, an evaluation instrument with high quality is an instrument which includes the most information into its analysis. It is costly to acquire information, of course. The principal looks at the relationship between quality and evaluation score and uses this relationship to calculate a score, which reveals the impact of undesired activities. Therefore, he uses the worse tests as information for manipulative behaviour.

To deter gaming completely, it is crucial that some manipulation cannot be manipulated. However, the proposed process allows better determent of manipulation than any single evaluation method, even if specific information is subject to specific manipulation.

The proposed mechanism and the underlying principal-agent model are in many aspects similar to the proposed weighting of multiple performance measures in Datar, Kulp and Lambert (DKL; 2001). Indeed, this paper provides a simplification of DKL to take account of notable differences. DKL and many other in the literature (e.g. Baker 2002) claim that incentive-compatible performance measurement implies a trade-off between risk and accuracy. However, the risk declines if a sufficient amount of relevant information is available at little cost. Adding information to a performance measure allows to identify the statistical relationship between information availability and measured performance. Think of added information as the next hour or day, and the tools from time-series econometrics can be applied to forecast the performance measure if another unit of information is added. This stepwise increase in the considered information also implies that more sophisticated evaluation methods are less risky, as more available information typically means a more precise estimation. As an additional difference to DKL, the model considers relationship of a principal with many agents with heterogeneous and unobservable abilities, which allows for self-selection processes among agents.

Though the mechanism and the underlying incentive contract are not dynamically modelled and the assumptions on the impact of gaming on different evaluation methods have been changed, the paper captures and enriches the results from CM. They show that an observable and manipulable output increases if it is included into an incentive contract, although the measure initially had a high correlation with the principal's objective. CM propose a lower incentive weight with increasing gaming and the introduction of a new measure if it implies lower impact of gaming. This paper proposes to keep a manipulable evaluation measure as an indicator for the extent of manipulation in a less manipulable evaluation measure.

Rewarding research quality may serve as an example. Research evaluators in Germany often use the number of publications as a proxy for quality, which is a cheap assessment method but dubious in terms of incentive compatibility (see Ursprung, 2003). Scientists are induced to publish countless papers with little if any new results if payments depend on such evaluations. Yet more papers are ceteris paribus better than less. Further measures such as citation impact, or the prestige of the journal could also be taken into account, for the price of higher evaluation costs. Manipulating these measures is difficult for most scientists. The proposed evaluation record would immediately reveal the excessive output of meaningless papers, as the evaluation score declines sharply with the additional measures.

The analysis in this paper is based on a simple incentive mechanism. The objective is to develop an evaluation mechanism which reduces waste and inefficiency from a particular form of assymetric information. This focus sets the paper apart from other literature on incentives in sectors like education, which look on non-pecuniary benefits (e.g. Besley and Ghatak, 2004), multi-tasking aspects (Holmstrom and Milgrom, 1991) or career concerns (e.g. Dewatripont, Jewitt, and Tirole, 2003). The argument in this paper takes into account that agents might have constraints which inhibit the implementation of the proposed evaluation mechanism. The paper analyses limited liability and risk-aversion in greater detail. Here, Kräkel (2004) provides a useful contribution to incentive schemes for agents with limited liability

The paper is structured as followed. The evaluation process and the proposed evaulation mechanism are introduced in the following section. Section 3 analyses the agency problem, at first in a benchmark case without manipulation and then in a case with manipulative evaluation results. Section 4 accounts for limited liability and section 5 for risk aversion. Section 6 tests the sensitivity of the results with respect to a change in the underlying assumptions.

2 The evaluation process

Evaluation constitutes an examination of the output of the agents in a specified period. Consider two methods, a good method T_g and a bad method T_b. The score of the good method reacts less sensitive to manipulation. The costs for this method are higher as well ($K_g > K_b$), the costs for a perfect evaluation are prohibitively high. More general, costs increase in quality. Both methods use the same score scale.

The time structure is the following. The principal chooses the evaluation mechanism and a wage contract. Then, the agents produce some output for one period. After the production, the output is evaluated and the payments are made. The principal evaluates honestly.

The principal can choose either method T_g, or method T_b, or a combination of both methods (T_c, with costs $K_c \geq K_g$). For simplicity, the evaluation score of agent i evaluated with method T is the sum of two linear functions and a method-specific error term.

$$Q_{i,T} = E_i + S_i(T) + \varepsilon_{i,T} \tag{1}$$

The two included functions are a desired-effort or "true output production" function

$$E_i = \theta_i e_i \tag{2}$$

and an undesired-effort, manipulation or "superficial output" function

$$S_i(T) = \gamma_T \mu_i s_i \tag{3}$$

The evaluation depends on characteristics of the agent and the properties of the evaluation method. The agent's ability in producing the desired output (e.g. his research ability) is denoted with θ_i and e_i is the desired effort (real input in research). Agents have also an ability to manipulate (μ_i) and can choose some undesired manipulation effort s_i. Manipulation thus constitutes an investment in superficial output, i.e. window-dressing activities. Both abilities θ and μ are distributed between 0 and 1 according to identical independent normal distributions, with $cov(\theta, \mu) = 0$ and the density function $g(\theta) = g(\mu)$. The individual abilities are private knowledge, their distributions are common knowledge. To get explicit solutions, the cost functions of the effort inputs are specified as

$$C(e_i) = \frac{e_i^2}{2}, e_i \geq 0 \tag{4}$$

$$C(s_i) = \frac{s_i^2}{2}, s_i \geq 0 \tag{5}$$

Section 6 discusses the impact of interdependencies between both cost functions.

For the parameter γ_T, assume the following. The good method, T_g, contains the information set G, the bad method T_b contains the information set B, with $B \cap G$. The intersection $B \cap G$ is not empty. Let $a > 0$ denote the share of B in G and give all units of information equal weight in the evaluation. All information in B can be manipulated with the parameter γ_b. Manipulation affects all units of information in B identically. Let the manipulation parameter for all in information in $G \ni B$ be 0, such that

$$a\gamma_b = \gamma_g \tag{6}$$

Hence, T_g is a more sophisticated version of T_b. Both test-scores are subject to manipulation.

The evaluation method has a two-fold impact on the observed quality. The better method T_g is less prone to superficial output than method T_b (i.e. $\gamma_g < \gamma_b$, with $\gamma = 0$ for a perfect test). The measurement error is subject to some evaluation-specific and student-specific observation error $\varepsilon_{i,T}$. Furthermore, the realisations of $\varepsilon_{i,T}$ are identically and independently distributed across students as well as across evaluation methods.

$$\varepsilon_T \sim N(0, \sigma_T^2) \text{ with cdf } F_T(\varepsilon_{iT})$$

The respective cumulative distribution function is denoted with $F(\varepsilon_{i,T})$ and the density function with $f(\varepsilon_{i,T})$. The iid assumption may not be the most realistic, but allows a simpler analysis of the second moment effects without loss of generality for the results.

2.1 The combined evaluation process

The principal and the agents know $0 < \gamma_g < \gamma_b$, i.e. to what extend each test is prone to manipulation. The principal calculates the combined evaluation score with the help of the function $G_c\left(Q_{i,1}, Q_{i,2}\right)$:

$$G_c\left(Q_{i,b}, Q_{i,g}\right) = Q_{i,c} = Q_{i,b} - \left(\frac{\gamma_b}{\gamma_b - \gamma_g}\right)\left(Q_{i,b} - Q_{i,g}\right) \tag{7}$$
$$= \left(\frac{\gamma_b}{\gamma_b - \gamma_g}\right) Q_{i,g} + \left(1 - \left(\frac{\gamma_b}{\gamma_b - \gamma_g}\right)\right) Q_{i,b}$$

The principal uses this function because it provides a consistent estimation of the "true" or desired output.

Result 1 *The expected desired output (E_i) is equal to the expected value of the combined evaluation.*

Proof. Notice that

$$E\left\{G_c\left(.\right)\right\} = \left(E_i + S_i(T_b) + \varepsilon_{ib}\right) - \left(\frac{\gamma_b}{\gamma_b - \gamma_g}\right)\left(E_i + S_i(T_b) + \varepsilon_{ib} - E_i - S_i(T_g) - \varepsilon_{ig}\right) \tag{8}$$

This equation can be transformed into

$$E\left\{G_c\left(.\right)\right\} = \left(E_i + \gamma_b \mu_i s_i\right) - \left(\frac{\gamma_b}{\gamma_b - \gamma_g}\right)\left(\gamma_b - \gamma_g\right)\mu_i s_i \tag{9}$$

which implies

$$E\left\{G_c\left(.\right)\right\} = E_i \tag{10}$$

■

As manipulation affects all information in B identically, the principal does not even have to include all information from T_b in T_g to deter gaming in T_c. He can run leaner estimations. This would hold even if information units in $G \ni B$ could be manipulated specifically, but only by pushing the measure upwards. Scholastic achievement tests provides an example for this caveat. Test specific manipulation will become detrimental if the test score is seen as a comparatively manipulative evaluation instrument and remains outside T_g, but inside T_b. Then schools could induce their tested students to score poorly in the test, which would lead to a high combined evaluation score.

The expected value of the combined error term ε_c is given by

$$\varepsilon_{ic} = \left(\frac{\gamma_b}{\gamma_b - \gamma_g}\right) E\left(\varepsilon_{ig}\right) + \left(1 - \left(\frac{\gamma_b}{\gamma_b - \gamma_g}\right)\right) E\left(\varepsilon_{ib}\right) = 0 \tag{11}$$

Hence, the properties of ε_{ic} are as follows:

$$\varepsilon_{ic} \sim N(0, V(\varepsilon_{ic})) \text{ with cdf } f_c(\varepsilon_{ic})$$

with

$$V(\varepsilon_{ic}) = \left(\frac{\gamma_b}{\gamma_b - \gamma_g}\right)^2 \sigma_g^2 + \left(1 - \left(\frac{\gamma_b}{\gamma_b - \gamma_g}\right)\right)^2 \sigma_b^2 \tag{12}$$

Result 2 *(i) Given that only two imperfect single evaluation methods exist, the variance of the combined error term is always larger as the variance of the best single evaluation method*

$$V(\varepsilon_c) > V(\varepsilon_g) > 0.$$

(ii) The variance of the combined evaluation converges to infinity if the available evaluation methods are almost equal with respect to quality

$$\lim_{\gamma_g \to \gamma_b} V(\varepsilon_{ic}) = \infty \tag{13}$$

and/or if both available evaluation methods are very bad.

$$\lim_{\gamma_g \to \infty} V(\varepsilon_{ic}) = \infty, \gamma_g < \gamma_b \tag{14}$$

Proof. The variance of method T_g is σ_g^2. The difference in evaluation quality $\gamma_b > \gamma_g > 0$ implies

$\left(\frac{\gamma_b}{\gamma_b - \gamma_g}\right) > 1$. Therefore, the first term in (12) is larger than σ_g^2 (i). Statement (ii) holds because of

$$\lim_{\gamma_g \to \gamma_b} \frac{\gamma_b}{\gamma_b - \gamma_g} = \infty \tag{15}$$

and

$$\lim_{\gamma_g \to \infty} \left(\frac{\gamma_b}{\gamma_b - \gamma_g} \mid \gamma_g < \gamma_b \right) = \infty \tag{16}$$

■

Result 1 states that, on average, the combined evaluation reveals manipulation. This revelation is caused by the negative weight for the less sophisticated test score. Result 2 shows that any principal applying the mechanism may face a trade-off between evaluation quality and risk if the possibility to improve a test are small $\left(\text{i.e. } \gamma_g \tilde{=} \gamma_b\right)$. However, the risk of the combined evaluation can be lower than the risk of method T_b if an evaluation method can be substantially improved and $\sigma_b^2 > \sigma_g^2$ holds. Taking additional information into account allows to hedge against shocks, which does not affect all information symmetrically. Then, the actual trade-off is between measurement quality and measurement cost.

Result 1 requires that two evaluations have an intersect of used information. In particular agents can invest in one type of unspecific manipulation only $(S_i(T_c) = \gamma_c \mu_i s_i)$. This manipulation affects all information in the bad test and some in the good test. The qualification is important because if agents could manipulate a single evaluation score specifically while keeping all others constant, the results from Result 1 could not be maintained.

3 The Principal-Agent-Problem

3.1 The benchmark case with no manipulation

To begin with, suppose that evaluation results cannot be manipulated by the agent (i.e. $\gamma_T = 0$, for T_g and T_b). Agent i invests in effort, produces an output and gets a wage w, with

$$w = \alpha + \beta \left(Q_{i,T} \left(\theta \right) \right) \tag{17}$$

The wage is composed of a fixed payment α and a variable pay factor β. The agent's utility function is

$$\max_{e} U_i = w - C(e_i) \tag{18}$$

with the respective first order condition:

$$\frac{\partial U}{\partial e} = \beta \theta_i - e_i^* = 0 \tag{19}$$

This first order condition (19) allows to derive the incentive compatibility constraint in the principal-agent-model

$$\beta \theta_i = e_i^* \tag{20}$$

The expression e_i^* denotes the effort level which is optimal for the agent. The principal wants to maximize his expected surplus. He achieves this objective by setting the wage components α and β as well as choosing the appropriate evaluation method. The principal can choose T_g, T_b, or their combination, T_c. Hence, his problem is

$$\max_{\alpha,\beta,T} \pi = \int_{\theta=0}^{1} \left[E\left(\theta \right) - K_T - \alpha - \beta \left(Q_{i,T} \left(\theta \right) \right) \right] d\theta, T \in \{T_g, T_b, T_c\} \tag{21}$$

For simplicity it is assumed that the total surplus accrues to the agents. Hence

$$\int_{\theta=0}^{1} \int_{\mu=0}^{1} \left[E\left(\theta \right) - K_T - \alpha - \beta \left(Q_{i,T} \left(\theta \right) \right) \right] d\theta d\mu = 0 \tag{22}$$

constitutes the participation constraint of the principal. Inserting from (22) into (21) yields

$$\max_{\alpha,\beta,T} \pi = \int_{\theta=0}^{1} \left[\theta e_i^* - K_T - \frac{(e_i^*)^2}{2} \right] d\theta \tag{23}$$

Using the incentive compatibility constraint derived from (19) the principal can set optimal incentives for the desired effort:

$$\frac{\partial \pi}{\partial \beta} = \int_{\theta=0}^{1} \theta^2 d\theta - \beta \int_{\theta=0}^{1} \theta^2 d\theta = 0 \tag{24}$$

which implies

$$\beta = 1 \tag{25}$$

Hence, the variable payment factor is identical for all agents whatever their ability is.

The fixed payment can be calculated with help of the principal's participation constraint (22):

$$\alpha = -K_T \tag{26}$$

Since the incentive mechanism is efficient the principal will always the evaluation method with the lowest cost.

The incentives under combined evaluation (T_c) are identical to those in this benchmark case, as the combined evaluation reveals manipulation completely (see Result 1). This revelation implies, in terms of manipulability, $\gamma_c = 0$, though with $\sigma_c^2 > \sigma_g^2 > 0$. The evaluation costs are higher as more information is required to observe the relationship between evaluation quality and measurement.

3.2 The case with manipulation of results

Now suppose $\gamma_T > 0$ for $T \in \{T_g, T_b\}$. Agent i invests in effort and manipulation to maximize their utility. Hence, the agent's objective function is given by

$$\max_{e,s} U_i = w - C(e_i) - C(s_i) \tag{27}$$

with

$$w = \alpha + \beta \left(Q_{i,T} \left(\theta, \gamma_T, \mu \right) \right) \tag{28}$$

and the respective first order conditions:

$$\frac{\partial U}{\partial e} = \beta\theta_i - e_i^{**} = 0 \tag{29}$$

$$\frac{\partial U}{\partial s} = \beta\gamma_T\mu_i - s_i^{**} = 0 \tag{30}$$

The first-order-conditions (29) and (30) allow to calculate the incentive compatibility constraints in the principal-agent-model, with e_i^{**} and s_i^{**} denoting the optimal effort and manipulation levels of the agent:

$$\beta\theta_i = e_i^{**} \tag{31}$$

$$\beta\gamma_T\mu_i = s_i^{**} \tag{32}$$

The ratio of both inputs depend on the respective abilities of the agent and the evaluation quality.

$$\frac{e_i^{**}}{s_i^{**}} = \frac{\theta_i}{\gamma_T\mu_i} \tag{33}$$

A change in the incentive factor β causes proportional changes of both inputs. Only a change in the evaluation quality changes the relative effort supply via γ_T. Furthermore, equation (30) reveals that perfect evaluation makes manipulation efforts prohibitively costly for the agent.

Result 3 *If the agents can manipulate the results, incentives are lower than in the case of no* *manipulation. The second-best outcome is a one-size-fits-all contract, i.e. every agent gets the same contract.*

Proof. Now, the principal's profit function is given by

$$\max_{\alpha,\beta,T} \pi = \int\limits_{\theta=0}^{1}\int\limits_{\mu=0}^{1} \left[E\left(\theta\right) - K_T - \alpha - \beta\left(Q_{i,T}\left(\theta,\mu\right)\right)\right] d\theta d\mu \tag{34}$$

Again, the principal can choose the evaluation method among T_g, T_b and T_c. He has to take the incentive compatibility constraints of his agents into account. The zero-profit condition for the principal is still holding (see (22)), which implies the following optimization problem

$$\max_{\alpha,\beta,T} \pi = \int\limits_{\theta=0}^{1}\int\limits_{\mu=0}^{1} \left[\theta e_i^{**} - K_T - \frac{\left(e_i^{**}\right)^2}{2} - \frac{\left(s_i^{**}\right)^2}{2}\right] d\theta d\mu \tag{35}$$

with the respective first order condition for β and T:

$$\frac{\partial \pi}{\partial \beta} = \int\limits_{\theta=0}^{1}\int\limits_{\mu=0}^{1} \left[\theta^2 - \beta^*\left(\theta^2 + \left(\gamma_{T^*}\mu_i\right)^2\right)\right] d\theta d\mu = 0 \tag{36}$$

$$\frac{\partial \pi}{\partial T} = -K_T^{'} - \frac{\partial \gamma_T\left(\gamma_T\left(\beta^*\right)^2\mu_i^2\right)}{\partial T} = 0 \tag{37}$$

The optimal incentive weight is given by β^*, the optimal evaluation quality by T^*. The first order conditions imply

$$\beta^* = \int\limits_{\theta=0}^{1}\int\limits_{\mu=0}^{1} \frac{\theta^2}{\theta^2 + \left(\gamma_T^*\mu_i\right)^2} d\theta d\mu < 1 \tag{38}$$

and an evalution quality where marginal costs and marginal benefits of evaluation quality are equal. Since perfect evaluation is prohibitively expensive, the evaluation cost function is convex, wehich ensures an interior solution. The marginal benefits of evaluation quality include an increase in the incentive weight and the direct decrease in manipulation. Recall that $\frac{\partial \gamma_T}{\partial T} < 0$.

The principal does not know the individual abilities and the agents cannot reveal their abilities through the choice of an individual incentive contract because all agents would choose a contract with $\beta > \iint \frac{\theta^2}{\theta^2+\left(\gamma_T^* \mu_i\right)^2} d\theta d\mu$ as $\frac{\partial U_i}{\partial \beta} > 0$ holds for all agents. ■

The fixed payment α is not necessarily negative anymore:

$$\alpha = -K_T + (1-\beta) E\left(Q_{i,T}(\theta, \mu)\right) \tag{39}$$

Note that $E\left(Q_{i,T}(\theta, \mu)\right)$ is identical for all agents, as the principal has no ex-ante information about the characteristics of an individual agent.

In summary, the principal sets lower incentives with increasing manipulability of the evaluation methods, otherwise the budget constraint will be violated. Hence, the principal will choose the combined evaluation method to eliminate manipulation if its costs are sufficiently low.

4 Limited Liability of the agents

In many cases, agents cannot bear losses indefinitely but require a minimum income. In this paper limited liability of the agents implies $w \geq 0$. No agent can afford a negative income. Limited liability restricts the ability of the principal to apply the combined evaluation.

Result 4 *With limited liability the incentives of the combined evaluation mechanism are weaker than in the case without. Suppose the costs of evaluation method b are zero, such that $K_c = K_g$. Then exists always an extrapolation function that ensures that the combined evaluation dominates evaluation method g.*

Proof. The fixed income in the benchmark case is negative in order to cover the costs of the evaluation (see (26)). The principal has to reduce the variable payment in order to increase the fixed payment (as in (39)), which will reduce the incentives. Furthermore, the result of a combined evaluation can be negative, too (see (7)). Hence, the principal has to restrict the possible results of the combined evaluation, which invites for manipulation (see Result 1). Previously, the principal used following function for the calculation of the combined quality score:

$$Q_{i,c} = \left(\frac{\gamma_b}{\gamma_b - \gamma_g}\right) Q_{i,g} + \left(1 - \left(\frac{\gamma_b}{\gamma_b - \gamma_g}\right)\right) Q_{i,b} \tag{40}$$

He has to change the function such that the second summand is smaller than the first one, e.g. by adding the following rule:

$$Q_{i,c} = 0 \text{ if } \left(\frac{\gamma_b}{\gamma_b - \gamma_g}\right) Q_{i,g} < \left(1 - \left(\frac{\gamma_b}{\gamma_b - \gamma_g}\right)\right) Q_{i,b}$$

In this case the relation $Q_{i,c} < Q_{i,g}$ holds if the score from method b is better than the one from method g. As the results from method b are more distorted by manipulation, $Q_{i,c}$ is closer to the desired output and less wasteful. ■

The first part of the result shows that limited liability restricts the possibility to eliminate manipulative inputs and induce desired effort. However, part two states that the combined evaluation is the most appropriate method if the principal modifies the extrapolation function (7). If combined evaluation becomes too costly, then the better single evaluation method dominates, of course.

5 Risk-aversion of the agents

The incentive effects of the combined evaluation changes again, if agents are risk-averse. The utility function remains as above

$$\max_{e,s,\gamma_T} U_i = w - C(e_i) - C(s_i) \tag{41}$$

However, the wage includes a compensation for the risk of the evaluation, the risk premium $RP(r, \beta^2 V(\varepsilon_T))$.

$$w = \alpha + \beta Q_{i,T}(\theta, \gamma_T, \mu) + RP(r, \beta^2 V(\varepsilon_T)) \tag{42}$$

This premium is a function of the size of the risky component of the payment, indicated by β. The component is risky because it depends on the evaluation score, which is associated with some uncertainty (ε_T). The degree of risk-aversion is given by r. It is assumed to be homogeneous among the agents. The risk premium increases in all arguments.

The participation and incentive compatibility constraints (22), (29) and (30) still hold.

Result 5 *Given a sufficiently high degree of risk-aversion among agents, combined evaluation will be dominated by the best single method (T_g).*

Proof. See appendix. ■

Unlike in the case of limited liability, risk-aversion not only modifies the application of combined evaluation method. Risk-aversion can actually inhibit it, because the combined evaluation method is associated with high risk. This high risk decreases the size of the variable payment (β).

Improving evaluation quality allows to decrease the risk of the combined evaluation. Yet, this does not necessarily make it the dominant solution, as the best single evaluation method improves as well. Adding more information decreases the risk of the combined evaluation more than the risk of applying only the best single, but manipulation is also better identified. However the principal can improve the situation by allowing for self-selection.

Result 6 *Let agents choose between the combined evaluation method T_c or the best single evaluation method T_g and let the principal set optimal incentive parameters for each method. Then*

1. *an agent with a high θ and a low μ will choose T_c;*
2. *an agent with a low θ and a high μ will choose T_g;*
3. *an agent will choose T_c if he was indifferent between T_g and T_c when evaluation methods were assigned by the principal;*
4. *the choice of evaluation methods by the agents dominates the assignment by the principal.*

Proof. First recall that the risk of an evaluation is constant for all agents regardless of their specific abilities. Then, the first two results derive from (27) and the resulting relationship between expected income and associated risk. For the third result, note that the agent was indifferent when all agents where evaluated in the same way. Due to the self-selection, the rewards of the best single method T_g decline, as all agents with a high productivity and little manipulation ability choose against it and the principal cannot differentiate between agents who game a lot and those who do less (see Result 1). Self selection dominates as incentives increase with productivity. ∎

With sufficiently risk-averse agents, the principal can take into account that people with different characteristics have a different perspective on a given risk. The "good" and "honest" agents have less to fear from the combined evaluation and will choose it eventually. The "average" agents will join them, because otherwise they will be lumped together with the not so good and honest agents. Because the principal cannot distinguish these agents, the "average" agents would end-up cross-subsidizing those "less desirable" fellows.

6 Variation of the model

In this section some of the underlying assumptions are relaxed to investigate how the results change with a modification of these assumptions.

6.1 Individual contracts

Evaluating the agents over time allows the principal to identify the abilities of the agents. Hence, he is able to design specific contracts for each agent. This does not modify the contract design when agents are risk neutral and the combined evaluation is applied. However, it affects the value of applying the most suitable single evaluation method.

Result 7 *The principal proposes individual contracts for each agent if he knows his abilities. With individual contracts, the combined evaluation becomes less attractive for agents with a high θ.*

Proof. For the contract design, see appendix.

Under a one-size-fits-all contract, the more able and less manipulating agents subsidized their counterparts. This transfer payment is eliminated with individual contracts, which in turn makes the best single method more attractive for the more able and less manipulating agents. ■

As a consequence from this result, the combined evaluation mechanism is more attractive, if the characteristics of the agents are unknown, because more productive agents are more discouraged with one-size-fits-all contracts than with individual contracts.

6.2 Interdependence between the cost functions

Previously, the costs of the different inputs were independent. However, marginal costs for one input may depend on the provision of the other input. Let $C_1(e_i, s_i)$ substitute $C(e_i)$ and replace $C(s_i)$ with $C_2(e_i, s_i)$ such that

$$\frac{\partial C_1(e_i, s_i)}{\partial e_i} = C_1' > 0; \frac{\partial^2 C_1(e_i, s_i)}{\partial (e_i)^2} = C_1'' > 0; \frac{\partial C_2(e_i, s_i)}{\partial s_i} = C_2' > 0; \frac{\partial^2 C_2(e_i, s_i)}{\partial (s_i)^2} = C_2'' > 0$$

hold.

Four relationships between the cost functions are possible

1. $\frac{\partial C_1'}{\partial s_i} < 0$
2. $\frac{\partial C_1'}{\partial s_i} > 0$
3. $\frac{\partial C_2'}{\partial e_i} < 0$
4. $\frac{\partial C_2'}{\partial e_i} > 0$

Only the first of these effects changes the case for the application of the combined evaluation mechanism qualitatively. If manipulation would increase costs for the desired input (case 2), it would strengthen the case for the combined mechanism. The third case ($\frac{\partial C_2'}{\partial e_i} < 0$) also increases the necessity to go again window-dressing. If the last case would hold, the entire discussion about gaming and manipulation would be obsolete.

Therefore, the focus is on the impact of $\frac{\partial C_1'}{\partial s_i} < 0$. Such a property changes the characteristics of what has been called manipulation. If more provision of this input decreases the costs for the desired input, then it is not entirely wasted. One may call the phenomenon advertising, which is in itself unproductive but generates a positive externality on the desired objective. To stay in the research example, redundant publications by a researcher may spread his results wider than a single paper and induce more people to use the results for further research. An other label for such a relationship may be a trial-and-error approach. People work superficially on a lot of projects. Any of these projects may carry the big idea, so the pursuit of many projects may reduce the search costs for finding that good idea.

For any realistic consideration, $\frac{\partial^2 C'(e_i)}{\partial s_i \partial e_i} > 0$ and $\frac{\partial^2 C'(e_i)}{\partial s_i^2} > 0$ have to hold, otherwise the desired effort could be increased without limits.

Result 8 *Suppose* $\frac{\partial C_1'}{\partial s_i} < 0$ *and*

$$\frac{\partial C_1'}{\partial s_i} = \frac{\partial C_2'}{\partial e_i} = \frac{\partial C_2'}{\partial e_i} = 0$$

hold. Assume that agents are risk-neutral. Then,

1. *the combined evaluation mechanism does not deter agents from showing some manipulative effort,*

2. *the evaluation quality from the best available single method is lower than in the case of* $\frac{\partial C_1'}{\partial s_i} = 0$ *and*

3. *the utility derived from this best available single method is greater than in the case of* $\frac{\partial C_1'}{\partial s_i} = 0$.

Proof. See appendix ■

Agents will always game, even with the combined evaluation process being applied, because gaming facilitates the production of the desired output. Principals have a lower desire to deter agents from gaming, since it has some positive side effects.

7 Conclusions

The paper has introduced an evaluation mechanism which eliminates incentives for manipulation even if individual methods cannot perfectly detect manipulation. The mechanism was based on the extrapolation of the perfect evaluation score from the scores of imperfect evaluations. Given full information about test quality and ability distributions, the "true" evaluation score can consistently be estimated. The implications of the argument are straight-forward. If there are several evaluation scores attached to a given output, it is not the best idea to take the arithmetic mean of all scores. The quality of the evaluation methods has to be considered, too. A declining score at increasing evaluation quality indicates that there is more window-dressing than real substance. Research evaluation is an area for a possible application. Here, methods with differing degrees of sophistication lead to different results for a given output and simple methods are easily manipulated.

The proposed mechanism sets better incentives than the application of any single evaluation mechanism if agents are not too risk-averse. If the liability of the agents is limited, combined evaluation may not completely deter from manipulation. If agents are risk-averse the selection of the evaluation method by each agent dominates uniform assignment by the principal. In this situation, the principal can improve his situation by allowing each agent to choose between an individual evaluation method and the combined mechanism.

The evaluation of research quality offered an example for application. Since the proposed mechanism is associated with a relatively large risk, it is more suitable for the evaluation of research organizations than individual scientists.

Appendix

Proof of result 5

The problem of the principal is given by

$$\max_{\alpha,\beta,T} \pi = \int_{\theta=0}^{1} \int_{\mu=0}^{1} \left[\theta e_i^{**} - \frac{\left(e_i^{**}\right)^2}{2} - \frac{\left(s_i^{**}\right)^2}{2} - RP\left(r, \beta^2 V(\varepsilon_T)\right)\right] d\theta d\mu \tag{43}$$

such that the following necessary conditions hold:

$$\frac{\partial \pi}{\partial \beta} = \int_{\theta=0}^{1} \int_{\mu=0}^{1} \left[\theta^2 - \beta^{**}\left(\theta^2 + \left(\gamma_{(T^{**})}\mu_i\right)^2\right)\right] d\theta d\mu - 2\beta RP'\left(r, V(\varepsilon_{(T^{**})})\right) = 0 \tag{44}$$

$$\frac{\partial \pi}{\partial T} = -K_T^{'} - \frac{\partial \gamma_T \left(\gamma_{T^{**}} \left(\beta^{**}\right)^2 \mu_i^2\right)}{\partial T} - \frac{\partial RP\left(r, \left(\beta^{**}\right)^2 V(\varepsilon_T)\right)}{\partial V(\varepsilon_T)} \frac{\partial V(\varepsilon_T)}{\partial T} = 0 \tag{45}$$

The optimal incentive weight for risk-averse agents is given by β^{**}, the optimal evaluation quality by T^{**}. The first condition allows to derive the variable payment factor:

$$\beta^{**} = \int_{\theta=0}^{1} \int_{\mu=0}^{1} \frac{\theta^2}{\theta^2 + \left(\gamma_{T^{**}}\mu_i\right)^2 + 2RP'\left(r, V(\varepsilon_{T^{**}})\right)} d\theta d\mu \tag{46}$$

Result 2 states that the variance of the combined evaluation is greater than the one of the best single method. Hence, the associated risk-premium is higher. This greater variance implies that the incentives from the best single evalution method T_g are greater than the incentives from combined evaluation for a sufficiently high r.

Proof of result 7

The contract for an agent i solves the following problem

$$\max_{\alpha,\beta_i,T} \theta_i e_i - K_T - \alpha - \beta\left(Q_{i,T}\left(\theta,\mu\right)\right)$$

The incentive compatibility constraints (31) and (32) hold as well as the zero-profit condition for the principal. With risk averse agents, the incentive weight is

$$\beta_i^{***} = \frac{\theta_i^2}{\theta_i^2 + 2RP'\left(r, V\left(\varepsilon_{T^{**}}\right)\right)}$$

for the combined evaluation and

$$\beta_i^{***} = \frac{\theta_i^2}{\theta_i^2 + \left(\gamma_T \mu_i\right)^2 + 2RP'\left(r, V\left(\varepsilon_{T^{**}}\right)\right)}$$

for an evaluation method T. Hence the incentive weight depends on the individual abilities. With risk neutrality, the contract associated with the combined evaluation is like in section 3.

Proof of result 8

The utility function for an agent i turns into

$$\max_{e,s} U_i = \alpha + \beta\left(Q_{i,T}\left(\theta\right)\right) - C_1(e_i, s_i) - C_2(e_i, s_i) \tag{47}$$

with the first order conditions

$$\frac{\partial U_i}{\partial e_i} = \beta\theta_i - C_1' - \frac{\partial C_2(e_i, s_i^{***})}{\partial e_i} = \beta\theta_i - C_1' = 0 \tag{48}$$

$$\frac{\partial U_i}{\partial s_i} = \beta\gamma_T\mu_i - C_2' - \frac{\partial C_1(e_i^{***}, s_i)}{\partial s_i} = 0 \tag{49}$$

Let s_i^{***} and e_i^{***} denote again the incentive compatible inputs. The incentive compatibility constraints imply

$$\beta = \frac{C_1'}{\theta_i} = \frac{C_2' + \frac{\partial C_1(e_i^{***}, s_i)}{\partial s_i}}{\gamma_T \mu_i} \tag{50}$$

Given his participation constraint,the principal solves the following problem:

$$\max_{e,s,T} \theta_i e_i - K_T - C_1(e_i, s_i) - C_2(e_i, s_i) \tag{51}$$

The first order conditions are

$$\theta_i - C_1' = 0 \tag{52}$$

$$\gamma_T \mu_i - C_2' - \frac{\partial C_1(e_i^{***}, s_i^{***})}{\partial s_i} = 0 \tag{53}$$

$$-K_T' - \frac{\partial \gamma_T \mu_i}{\partial T} = 0 \tag{54}$$

The second of these conditions implies $s_i^{***} > 0$ even for $\gamma_T = 0$. (1. statement).

Integrating conditions 53 and 54 yields

$$K_T' = \frac{\partial \gamma_T \mu_i}{\partial T} \left(1 - \frac{\partial C_2'}{\partial \gamma_T \mu_i} - \frac{\partial C_1(e_i^{***}, s_i^{***})}{\partial \gamma_T \mu_i} \right)$$

The term $\frac{\partial C_1(e_i^{***}, s_i^{***})}{\partial \gamma_T \mu_i}$ reflects an additional marginal cost from increasing the evaluation quality. Hence, the quality of the best available single method is lower than in the previous sections (Statement 2).

The benefit of this best single method is higher because with $\frac{\partial C_1(e_i^{***}, s_i^{***})}{\partial \gamma_T \mu_i} = 0$ the evaluation quality of the most suitable single method would be higher. Now however, a method with less quality dominates any more precise measurement bar the combined evaluation process. (statement 3)

References

Baker, G.; Gibbons, R. und Murphy, K. J. (1994), "Subjective Performance Measures in Optimal Incentive Contracts." The Quarterly Journal of Economics, 109(4), pp. 1125-56.

Baker, George (2002), "Distortion and Risk in Optimal Incentive Contracts." Journal of Human Resources, 37(4), pp. 728-51.

Besley, T. and Ghatak, M. (2004), Competition and Incentives with Motivated Agents, CEPR Discussion Paper No. 4641.

Courty, Pascal and Marschke, Gerald (2003), "Dynamics of Performance Measurement Systems." Oxford Review of Economic Policy, 19(2), pp. 268-84.

Datar, Srikant; Cohen Kulp, Susan and Lambert, Richard A. (2001) "Balancing Performance Measures." Journal of Accounting Research, 39(1), pp. 75-92.

Demougin, D. and C. Fluet (2003), Inequity Aversion in Tournaments, mimeo.

Dewatripont, M., Jewitt, I., Tirole, J. (1999), The Economics of Career Concerns, Review of Economic Studies, 66(1), 189-217.

Holmstrom, B. and Milgrom P. (1991), Multi-task Principal-Agent Analysis: Incentive Contracts, Asset Ownership, and Job-Design, Journal of Law, Economics and Organization 7 (special issue), 24-52.

Kerr, Steven (1975). "On the Folly of Rewarding for a While Hoping for B." Academy of Management Journal, 18(4), pp. 769-83.

Kräkel M. (2004), Tournaments versus Piece Rates under Limited Liability, mimeo.

Murphy, Kevin J. and Oyer, Paul (2003). "Discretion in Executive Incentive Contracts" Marshall School of Business, University of Southern California Graduate School of Business, Stanford University.

Ursprung, H. W. (2003), "Schneewittchen im Land der Klapperschlangen: Evaluation eines Evaluators", Perspektiven der Wirtschaftspolitik, Band 2, 177-189.