THE PERFORMANCE MEASUREMENT TRAP*

Dmitri Kuksov
(University of Texas at Dallas)

J. Miguel Villas-Boas
(University of California, Berkeley)

June, 2017

* Preliminary version. Comments are very welcome. Email addresses: dmitri.kuksov@utdallas.edu and villas@haas.berkeley.edu.
The Performance Measurement Trap

Abstract

This paper investigates the effect of performance measurement on the optimal effort allocation by employees, when firms are concerned about retention of employees with higher abilities. It shows that introducing employee performance measurement may result in productivity, profit, and welfare losses when all market participants optimally respond to the expected information provided by the measurement and the (ex-post) optimal retention efforts of the employer cannot be (ex-ante) contractually prohibited. In other words, the dynamic inconsistency of the management problems of inducing the desired effort allocation by the employee(s) and the subsequent employer’s objective to retain high ability employees may result in performance measurement yielding an inferior outcome.

Keywords: Game Theory, Contract Design, Principal-Agent problem, Salesforce Compensation
Oftentimes, the management of an organization is concerned with multiple objectives when creating an incentive structure for employees. Broadly speaking, one can divide them into the objectives linked to inducing the desired performance, and those objectives linked to hiring and/or retaining the most productive employees. The former include the objectives to induce employees to exert the optimal effort level overall and to allocate this effort optimally across the activities contributing to overall productivity. These are already non-trivial tasks when effort and effort allocation cannot be observed directly; finding the optimal incentives’ design for this set of objectives may be especially difficult when employees are heterogeneous. If employees are heterogeneous in a vertical dimension (ability), the second set of objectives (employee hire and retention policy) is also important and non-trivial. As is the case with effort allocation, the employees’ abilities are also usually not directly observable. Various performance measurements are commonly used to evaluate an employee’s ability for subsequent retention efforts as well as to incentivize the employees’ effort level and its allocation between productive activities.

The optimal design of the compensation packages (contracts) offered to employees involves providing the right incentives for the effort level and effort allocation choices of the employees subject to the individual rationality constraints (the value of the contract should not be worse than the value of the employee’s outside option). Naturally, the individual rationality constraints may also be unknown to management. This means that from the management’s point of view, the higher the value of the contract to an employee, the less likely the employee is to leave the organization. Usually this implies that the higher the employee’s ability, the better the contract parameters that management should offer to the employee.\footnote{When the contract rewards performance, and performance is correlated with ability, the same contract has a higher value to an employee with a higher ability. Therefore, contracts with the same parameters may be more likely to retain employees with higher abilities. However, one can empirically observe that it is usual for the parameters of the contract offered to an employee with better estimated ability (historical performance) to}
In a multi-period employment market where relatively short-lived contracts are renewed periodically (e.g., each year), the effort, effort allocation inducement, and employee retention problems are inter-temporally linked. Performance measurement in one period is used for current compensation according to the current incentive contract, but it also affects the future contract(s) offered to the employee due to the management’s (future) inference of the employee’s ability from the past (at the time of the inference) performance measurements. This in turn means that when deciding on the effort and its allocation, an employee needs to take into account not only the current contract’s incentives, but the expected effect of his effort and its allocation on the value he expects from the future contracts. In other words, the full incentives will always have an implicit component coming from the value of the future contracts.

Furthermore, management decides not only how to use the available information (performance measurements) about the employees, but it also decides whether and how to implement performance measurement as well (i.e., which kind of information about employees to obtain). The number of measures, or the amount of monitoring, is a frequently considered decision. The obvious trade-off is between the costs of measurement (e.g., the employees’ compliance costs) and the assumed benefit of the more extensive measurement which could lead, supposedly, to a better-designed incentive structure. It is well known that there could be a tendency to put too much weight on visible measures. One may hope that a rational manager would be able to optimally weigh each metric, so that the above problem is resolved, and more measurement would end up being beneficial to management. In other words, if management is fully rational, it seems intuitive that it should be able to design a better contract when more measurements are available. After all, it is within the power of management to not reward any of the metrics (see e.g., the flat wage result in Holmstrom and Milgrom, 1991).

However, the implicit nature of incentives discussed above raises the question as to whether management can commit to appropriately weigh metrics over the long haul. The contract re-also be better. For example, employees with superior past records or qualifications are often offered higher base salaries. This outcome is not unexpected: designing a contract to be more valued by a high-ability employee is akin to a second-degree price discrimination. Offering better contract parameters to employees with superior past records is akin to a third-degree price discrimination. One could expect a combination of both of these to be optimal.
newly introduces a dynamic inconsistency problem of the current objective of incentivizing effort and its allocation and the future objective of retaining employees whose record suggests better abilities. Since employees are forward-looking and can expect this inconsistency, introduction of a new metric results not only in an enlargement of the set of the available instruments for the management, but also in a change of the employees’ environment. While the larger set of instruments could help to achieve the management’s objectives and increase social welfare, the dynamic inconsistency could be detrimental. The net effect is not immediately clear.

In this paper, we show that the net effect could be negative by formalizing the following intuition. Suppose that the total effort a particular employee would exert is essentially fixed (i.e., not easily changed by incentives), and that in the absence of additional performance measurement(s), employees are incentivized to allocate effort across two components in an efficient way. However, suppose the existing incentive to allocate effort efficiently (using the existing measurements) is not particularly strong, so that an employee would distort his effort allocation if a new, non-negligible incentive to do so is introduced. For example, management may be able to observe and enforce the time spent at work, and any allocation of time across activities may not impose a disutility on the employees so far as they have to spend that time at work. In this case, a slight weight on overall output of the organization, even if this output is only weakly correlated with the effort allocation of a given employee, is sufficient to induce the employee’s optimal effort allocation. As a result, the system is working nearly perfectly. Different employees contribute differently to the total output due to their different abilities, but each does what he can (providing the effort he is capable of and efficiently allocating it across the effort components), and nothing can be done about the abilities themselves. In order to simplify the presentation, we will identify the total effort the employee is capable of with the employee’s ability. Now suppose that due to a random value of the outside option (e.g., due to the changing preferences for living close to the current employment area), each employee may leave employment at the beginning of each period, but this decision may, at least for some employees, be affected by the compensation that he is offered. Then, to encourage the better employees not to leave the organization

---

2Even if long-term contracts are possible, unless management can commit not to give out bonuses (which are not fully spelled out in the contract), the implicit nature of incentives remains present.
through an appropriate retention policy, i.e., by offering a better next period contract to the better employees, management is interested in evaluating its employees’ abilities.

Suppose further that the feasible measurement under consideration is the measurement of one, but not all, of the effort components. Note that without this additional performance measurement, given the optimal effort allocation, the employee’s ability (total effort) is perfectly correlated with each of the effort components. Then an employee’s ability could be perfectly judged from either component. This seems to provide a compelling argument for introduction of the performance measurement, even though it only measures one of the effort/performance components. However, knowing that management will offer a better contract in the next period to those employees who performed better on the measured component, if management starts measuring one of the dimensions of effort, and even if no weight is placed on it in the current compensation schedule, employees will distort effort allocation toward that dimension in hopes of securing a better contract offer in the next period. If this distortion is not costly for the employees, all effort may be allocated to only one component, thus, possibly having a negative effect on the total output.

But would not management then be able to adjust the current compensation package to eliminate this distortion, perhaps through a negative weight in the current contract on the performance component measured? The answer is: generally, no. The reason, again, is employee heterogeneity. Generally, one might expect that employees would have some independent and private estimate of their likelihood of staying with the organization next period. Therefore, the best management can hope to achieve with the optimally designed “reverse incentives” contract is that some employees, knowing that they are more likely than average to leave, will be incentivized by the reverse incentive to distort the effort towards the other component, while most of the remaining employees, having a higher likelihood to stay than management estimates, would still distort their effort toward the measured component. The outcome is that, while any aggregate mix of effort allocations can be achieved, almost all individuals may end up allocating effort inefficiently.

Note that perfect information, if costlessly available to management, should be welfare en-
hancing. However, imperfect measurement, insofar as it measures some, but not all, of the components of productive output, is likely to distort effort allocation. While this measurement can be used for better retention and a better total effort incentive policy, these improvements are at the expense of the efficiency of the effort allocation. Furthermore, inability to commit to future contracts means that this asymmetric measurement problem may not be feasible to fully resolve.\(^3\) Therefore, when deciding on whether to introduce a partial (or imperfect) measurement (i.e., a measurement that weights some of the effort/output components heavier than others), the management needs to decide whether the problems of the effort enforcement and employee retention are more important than the problem of effort allocation. Thus, the common practice to “collect the data first; decide what to do with it later” could be a treacherous path even if the management is fully rational and benevolent.

The remainder of the paper is organized as follows. The next section discusses the related literature. Section 3 presents the model, and Section 4 considers the cases when measurement is not possible, and when measurement is available on the first period effort level. Section 5 presents the effect of measurement being also possible on the second period effort choice, and Section 6 concludes. The proofs are collected in the Appendix.

2. RELATED LITERATURE

This paper builds on the extensive literature on the principal-agent and salesforce compensation problems.\(^4\) This literature, in particular, explores the optimal weights the principal needs to place on measurements to account for the effort distortion across multiple tasks (e.g., Holmstrom and

\(^3\) Management can offer long-term contracts, but the problem is that once management sees evidence of high ability and given the probability that the employee quits, it would then offer retention bonuses to some workers. That is to say, a commitment to never use the measurement is not renegotiation-proof. One could also envision “slavery contracts” that commit the employee to work forever. Although such contracts would solve the issue of commitment (since retention is no longer an issue), they would result in inefficiency since due to the random outside options (if they are not perfectly predictable by the employees), sometimes it is efficient for an employee to leave. From a profitability standpoint, such contracts would come at the expense of offering a higher base salary up front. Thus, the negative value of measurement would persist even if slavery contracts are allowed.

\(^4\) See, for example, Basu et al. (1985), Rao (1990), Raju and Srinivasan (1996).
For example, Holmstrom and Milgrom (1991) show that when the effort distortion between tasks is sufficiently severe, flat pay (i.e., contracts ignoring performance measurement) may be optimal, and discuss how it applies to the ongoing teacher compensation debate. Hauser et al. (1994) explore making the incentive scheme based on customer satisfaction measures as a way for employees to put effort on dimensions that have longer term implications. Bond and Gomes (2009) explore the ability of the principal to affect a multi-tasking agent’s efforts. The literature has also discussed what happens in a dynamic setting with the question of desirability of long-term contracts (e.g., Malcolmson and Spinnewyn, 1988) and renegotiation-proofness. One usual assertion is that the principal and welfare are not hurt by more information (measurement), but this is due to the assumed ability of the principal to put a sufficiently low weight on what is being measured. Alternatively, we consider the problem of the principal’s possible inability to commit to the future contracts and show that the optimal contract in the presence of more information (more performance measures) may lead to decreased profitability and social welfare.

Cremer (1995) considers a dynamic consistency problem of the principal’s commitment to incentivize effort by the threat of firing the agent. Assuming that the agent values job renewal, Cremer shows that in some cases, the principal would like to commit not to observe information about the reason the agent fails in order to increase the agent’s incentive to perform. By not observing the reason for failure and therefore effectively committing to fire the agent if output is not up to a standard, the principal is able to circumvent the positive-pay restriction and increase the incentive for the agent to perform well without increasing the expected pay. In contrast to Cremer, we examine the opposite problem: when the principal would like to commit not to incentivize the agent.

Dynamic (in)consistency issues have been studied in various other contexts, such as a durable-goods monopoly setting prices over time (e.g., Coase 1972, Desai and Purohit 1998), and central bank policies on money supply (e.g., Kydland and Prescott, 1977). In this paper, we essentially

---

5Some measurement can also be obtained on the market conditions by allowing lobbying by sales people on the incentive scheme (e.g., Simester and Zhang, 2014).
consider the implications of the dynamic inconsistency issue in a principal–agent framework.

3. The Model

Consider a principal–agent model with agents having different abilities and interacting with the firm (principal) during two periods, indexed by \( t = 1, 2 \). Since we abstract away from any effects of one agent’s behavior on other agents or the incentives the principal has in treating other agents, we consider, without loss of generality, a single agent whose type is uncertain to the principal. The principal’s (management’s) objective is to maximize the total expected payoff (profit) net of wage paid to the employee across the two periods by choosing a contract (wage conditional on observables) to offer the employee at the beginning of each period. Let \( \pi_t \) denote the profit gross of the expenditure on wages in period \( t \), and let \( C_t \) be the wage paid to the employee in period \( t \). To simplify the presentation, assume no discounting. Then, the firm’s problem is

\[
\max_{C_{1,2} \geq 0} \sum_{t=1}^{2} E(\pi_t - C_t)
\]  

(1)

The wage (contract) could depend on everything the principal observes prior to the offer or at the time of payment, as we assume the payment is done after the relevant period is over. However, the contract is restricted to provide non-negative pay to the employee for any outcome (limited liability). This could be justified, for example, because obtaining money from an employee who received no income in the current period may not be possible.

Period \( t \)'s profit is uncertain and depends on the employee's choice of the effort allocation \( \pi \equiv (x, y) \in \mathbb{R}^{+2} \) where \( x \) and \( y \) are the two dimensions of effort. To be clear, we will call the vector \( \pi \equiv (x, y) \) the effort allocation, and the sum of effort components \( e \equiv x + y \) the total effort. Assume the employee has a per-period budget constraint on his total effort \( x_t + y_t \leq e_i \), where \( e_i \) depends on the employee's type, which can be high, \( h \), or low, \( \ell \), that is, \( i \in \{\ell, h\} \). Assume that besides the budget constraint on effort, the employee has no intrinsic disutility of effort.

To simplify the potential contract structures, we model profit as possibly attaining one of just
two possible values, one of which is normalized to 1 and the other is denoted by $-b$, with $b > 0$, and model the effect of the employee's effort allocation as affecting the probability of achieving the high profit level. Assume that the probability of high profit increases in

$$a(x, y) = x \cdot y.$$  \hspace{1cm} (2)

This probability is maximized at $x = y = e_i/2$, i.e., the efficient effort allocation, which is also the most desired by the firm, is for the employee to equally split the total effort between the two effort components. For further simplification, we define the gross of wages profit in such a way that the expected profit equals $a(x, y)$ and, hence, refer to $a$ as the employee’s productivity. We denote the productivity’s maximal level by $\tilde{a}_i \equiv (e_i/2)^2$ and call it the ability of the type-$i$ employee. In other words, the employee’s productivity is his actual expected contribution to profits (gross of wages), while his ability is how much he can technically contribute if he were to maximize this contribution. Thus, the employee’s ability is a characteristic of the employee, while his productivity is his choice variable constrained by his total ability.

The above assumptions on the expected profit imply the following gross profit function specification as a function of the employee’s productivity:

$$\pi_t(a_{it}) = \begin{cases} 
1, & \text{with probability } (b + a_{it})/(b + 1); \\
-b, & \text{with probability } (1 - a_{it})/(b + 1),
\end{cases} \hspace{1cm} (3)$$

where $a_{it}$ is the employee’s choice of productivity in period $t$ (as defined by his effort allocation through Equation (2)), and $b > 0$ is a parameter. One can easily see that in this specification $E\pi_t = a_{it}$, but the informativeness of the profit realization (whether it is “1” or “$-b$”) about the employee’s actions depends on $b$ and tends to zero when $b$ tends to infinity. For clearer presentation, we focus the analysis on the case of large $b$, but discuss, when appropriate, what happens when $b$ is not too large. The case of large $b$ can be seen as a case in which the positive output is highly likely, but the loss is important when the positive output does not occur. In terms of the model, large $b$ means that the output obtained is not very informative of the employee’s ability.
To satisfy the constraint that the probabilities of the two profit outcomes are positive, we need $0 \leq e_\ell < e_h \leq 2$. (If $e_h = 2$, the efficient allocation $x = y = 1$ leads to $a_h = 1$.) To further simplify, assume $e_\ell = 0$ and $e_h = 2$, so that $\tilde{a}_\ell = 0$ and $\tilde{a}_h = 1$. Note that given $\tilde{a}_\ell = 0$, the allocation decision of the low type (allocating 0 across the two components) is immaterial. When $\tilde{a}_\ell = 0$, the probability of the profit being high (being "1") is $b/(b+1)$ and the probability of the profit being low (being "−b") is $1/(b+1)$. When $\tilde{a}_h = 1$ and the agent chooses the efficient effort allocation, the firm gets the high profit with probability one.

Note that if the agent is choosing the efficient effort allocation, a low profit implies that the agent has low ability for sure, but a high profit does not necessarily imply that the agent has the high ability. Note also that the greater $b$ is, the less informative the high profit will be of the employee’s ability.

Although we assume that an employee does not have a disutility of effort in any combination $(x, y)$ such that $x + y \leq e_i$, where $i$ is the employee’s type, the employee has an outside option $z$ in the second period, such that the employee may not accept the second period contract. This outside option could come from the utility of working elsewhere or the second period cost of traveling to work. This outside option allows us to consider the firm’s retention problem in the second period. For simplicity, we assume that $z$ does not depend on the employee’s type $i$. The outside option is zero in the first period so the employee accepts any non-negative offer in the first period.\(^6\)

To be more specific, assume that the second period’s outside option $z$ is uniformly distributed on $[0, 1]$. This outside option is known to the employee before his decision of whether to accept the second period offer, but may or may not be known to him before his choice of the effort allocation $\bar{z}_1$ in the first period. We first assume that the employee does not know $z$ before the effort-allocation stage of the first period; he learns it just before he needs to decide whether to accept the second period’s offer. We also later consider variations of the model to see how the results change if the employee can condition his effort allocation $\bar{z}_1$ on the second period outside

---

\(^6\)To avoid potential incentives for employees to signal high ability through refusing a first period offer in order to gain a higher wage in the second period, assume that the firm does not engage in hiring at all in the second period. From the equilibrium that we derive, such signaling will also not be optimal in this setting.
option $z$ or if he has some (but not full) information about it.

In addition to choosing the contract to offer, the principal needs to decide whether to introduce measurement $m(x)$ of the $x$-component of the employee’s effort. We assume that the measurement of the $y$-component, beyond its inference from the profit realization and any available measurement of $x$, is not feasible. In order to reduce the complexity of the compensation structure, assume that $m(x)$ may take only one of two values, call them 0 and 1, but the probability of the high measurement realization (“1”) increases in $x$. For analytical tractability assuming linearity and that the high type can ensure $m(x) = 1$ by allocating all effort toward $x$, we arrive at the following measurement specification:

$$m(x) = \begin{cases} 
1, & \text{with probability } x/e_h = x/2; \\
0, & \text{with probability } 1 - x/e_h = 1 - x/2.
\end{cases} \quad (4)$$

Note that with this measurement technology and given the assumptions $e_\ell = 0$ and $e_h = 2$, a measurement of “1” indicates that the agent has high ability, while a measurement of “0” does not necessarily indicate that the agent has low ability as long as the high ability agent is not expected to shift all his effort toward component $x$.

If measurement is introduced, the compensation in each period is a function of the current and past profits and measures of $x$. If measurement is not introduced, the compensation can only be a function of current and past profits. There are three conceptually different possibilities of the timing when $m(x)$ is introduced: (1) after the employee decided on his effort allocation $\bar{e}_1$ in the first period; (2) before the employee decided on his effort allocation $\bar{e}_1$ in the first period, but after the first period compensation rule was set (and it was done without having the possibility of measurement in mind), and (3) before the employee decided on his effort allocation $\bar{e}_1$ in the first period, and the first period compensation rule was determined with the possibility of the measurement in mind. We consider each of these cases. In the next section, we consider the case in which the measurement of the component $x$ is only possible in the first period effort. In Section 5, we consider the case in which the measurement of component $x$ is possible in both the first and second periods.
4. Model Analysis

In this section, we first derive the optimal compensation schedule in the benchmark case when the measurement of \(x\) is not a possibility. We then consider how the second period’s compensation and outcomes are affected if the measurement of the first period’s component \(x\) of the effort is unexpectedly introduced in period 1 after the employee already committed to his first period’s effort allocation. Next, we consider how the outcomes change if the employee optimally responds to the presence of the measurement of the first period effort, but assuming that the first period’s compensation rule is unchanged after the measurement is unexpectedly introduced. Conceptually, the last consideration is when the measurement introduction was unexpected by the firm, while the earlier one is when it was unexpected by the employee. Finally, we consider optimal compensation contracts in both periods given that the measurement of the first period’s \(x\) component is present, or is expected to be introduced. Comparing the outcomes under the last scenario with the previous ones, we show that the introduction of measurement could be detrimental to the principal and to social welfare. The case of measurement of \(x\) component in both periods is presented in Section 5.

4.1. Benchmark Case: Optimal Contract without Measurement

Given that the employee’s outside option is zero in the first period and that he does not have an intrinsic preference of how to allocate the maximum effort \(e_i\) at his disposal, compensation \(C_1 = \varepsilon 1_{\pi_1 = 1}\) with \(\varepsilon > 0\) achieves efficient allocation (so far as the employee does not expect the second period contract to provide perverse incentives for the employee to lower the firm’s expectation of their abilities).\(^7\) Thus, in equilibrium, we have efficient effort allocation and \(C_1 = 0\), i.e., the first-best for the principal (as contracts are not allowed to have negative pay).

In the second period, the problem is slightly more complicated. As the outside option is now positive and uniformly distributed on \([0, 1]\), the employee has a positive chance of leaving and a

\(^7\)The term \(1_{\text{condition}}\) is the indicator functions and takes the value of 1 if the “condition” is true, and takes the value of 0 otherwise.
positive chance of staying given any reasonable offer.\textsuperscript{8} Consider the second period’s contract

\[ C_2 = c_1 + c_2 1_{\pi_2=1}, \]  

(5)

where \( c_1 \geq 0 \) and \( c_2 \geq 0 \) are the firm’s decision variables and are functions of the observables after the first period (i.e., functions of \( \pi_1 \)). Thus in this benchmark case \( c_1 \) and \( c_2 \) may be functions of the first period’s profit realization and are non-negative to ensure a non-negative wage in any possible outcome.\textsuperscript{9}

Given this contract, the expected pay of the employee with ability \( \tilde{a}_i \) is

\[ c_1 + c_2 (b + \tilde{a}_i)/(1 + b) \]

if the employee chooses the efficient effort allocation, which is assured by \( c_2 > 0 \). Since the outside option is uniformly distributed on \([0, 1]\), this expected pay is also the probability that the employee stays. The expected net second period profit from the type-\( i \) employee who is staying and allocating his effort optimally for the firm is \( \tilde{a}_i - c_1 - c_2 (b + \tilde{a}_i)/(1 + b) \). Therefore, if the probability that the employee is of high type is \( P_h \), the expected net profit is

\[ E(\pi_2 - C_2) = (1 - c_1 - c_2) (c_1 + c_2) P_h - \left( c_1 + c_2 \frac{b}{1+b} \right) \left( c_1 + c_2 \frac{b}{1+b} \right) (1 - P_h). \]  

(6)

Note that \( P_h \) is the posterior probability of the employee having the high ability given the observables after the first period, i.e., the first period profit in this case. Maximizing the expected profit with respect to \( c_1 \) and \( c_2 \) under the constraint that \( c_1 \geq 0 \) (since a negative wage is not allowed), we find that the optimal second period’s contract has

\[ c_1 = 0 \quad \text{and} \quad c_2 = \frac{P_h (1+b)^2}{2(b^2 + P_h + 2P_h b)}. \]  

(7)

Note that as \( b \) increases from zero to infinity, the weight \( c_2 \) on the profit decreases from half of

\textsuperscript{8}Since profit is either one or negative, the optimal contract cannot provide expected utility to the employee that exceeds the outside option with probability one.

\textsuperscript{9}Although \( c_2 < 0 \) is technically allowed as far as \( c_1 + c_2 \geq 0 \), it is straightforward to check that it cannot be a part of an optimal contract – it would give an incentive to inefficiently allocate effort without the benefit of higher retention of the high type employees.
the high type ability (1/2) to half the expected ability of the employee (\(P_h/2\)). This is because as \(b\) increases, the profit realization becomes less informative of the employee’s ability. Therefore, given the profit-based-only incentive \((c_1 = 0)\), for low \(b\), the contract gives a low expected payoff to a low-ability employee and therefore, only the high-ability employee accepts the contract with a reasonable probability. But for high \(b\), contract acceptance becomes less dependent on the employee’s ability. Note also that if \(P_h = 1\), then the firm is indifferent between any \(c_1 > 0\) and \(c_2 > 0\) as far as their sum is the same; this is because \(\pi_2 = 1\) for sure for the high type. But if \(P_h < 1\), the above solution is uniquely optimal.

By differentiating \(c_2\) with respect to \(P_h\), one can see that the optimal offer increases in the principal’s belief about the employee’s ability:

\[
\frac{dc_2}{dP_h} = \frac{(1 + b)^2b^2}{2(b^2 + P_h + 2P_h b)^2} > 0. \tag{8}
\]

This means that if there is some way for the employee to demonstrate high ability without incurring a significant cost, he would strictly prefer to do so. This is the key to the result that a performance measurement of one effort component would be used by the high type employee to convince the employer that he is of high type at a cost of the first period profit. Of course, if profit is itself very informative or if the weight on the first period profit is sufficiently high, then this first period possible distortion by the employee may not happen. Also note that the above inequality means that, effectively, the expectation of the second period’s contract puts an implicit positive weight on the first period’s profit (since all else being equal, \(\pi_1 = 1\) implies a higher probability of the high type than \(\pi_1 = 0\) does), which gives the employee a strictly positive incentive to allocate the first period’s effort correctly. That is, even though the first period’s equilibrium contract leaves the employee indifferent as to how to allocate his effort if the employee were myopic, the equilibrium is actually strict due to the expected-by-the-employee second period contract’s positive dependence on the first period’s profit realization.

To complete the derivation of the equilibrium outcomes in the benchmark case, when the performance measurement is absent, it remains to derive the posterior probability \(P_h\) that the
employee is of high type as a function of the first period’s profit. Rational expectations also implies that this is what the firm expects). By Bayes’ rule:

\[ \Pr(h | \pi_1 = 1) = \frac{1 + b}{1 + 2b}, \quad \text{and} \quad \Pr(h | \pi_1 = -b) = 0. \]  

(9)

Substituting these in Equation (7), we obtain the second period payment \( C_2 \) to the employee as a function of the first and second period profits:

\[ C_2 = \hat{c}_2 \mathbf{1}_{\pi_1 = \pi_2 = 1}, \quad \text{where} \quad \hat{c}_2 \equiv \frac{(1 + b)^3}{2(1 + 2b)(1 + b + b^2)}. \]  

(10)

The expected net-of-wages profit is 1/2 in the first period (it is equal to 1 if the employee is of high ability and 0 if he is of low ability) and

\[ \mathbb{E}(\pi_2 - C_2) = \frac{(1 + b)^3}{8(1 + 2b)(1 + b + b^2)}. \]  

(11)

in the second period. This follows from substituting the optimal contract conditional on \( \pi_1 = 1 \), equation (10), into the expected profit, equation (6), and multiplying it by the probability that \( \pi_1 = 1 \), which is \( \frac{2 + b}{2 + 26} \). Note that the eventuality of \( \pi_1 = -b \) does not enter because it implies the employee is of low type and therefore, results in \( C_2 = 0 \) and no retention.

For example, for \( b = 0 \), we have \( C_2 = 1/2 \cdot 1_{\pi_1 = \pi_2 = 1}, \mathbb{E}(\pi_2 - C_2) = 1/8 \), and the expected total net profit of 5/8.\(^{10} \) For \( b = 10 \), we have \( C_2 = 0.29 \cdot 1_{\pi_1 = \pi_2 = 1}, \mathbb{E}(\pi_2 - C_2) = 0.07 \), and the expected total net profit of 0.57. As \( b \to \infty \), we have \( C_2 \to 1/4, \mathbb{E}(\pi_2 - C_2) \to 1/16 \), and the expected total net profit converging to 9/16.

Intuitively, the second period’s and the total net profits decline with \( b \) as the first period

\(^{10} \)Note that the assumption of relatively low retention rate (relatively high outside option) forces the profit to be low in the second period – both due to the low retention probability and the extra expenditure on wages. However, reducing the upper bound on the outside option would complicate the analysis as it would require considering the boundary case of the retention probability equal to one under a potentially optimal contract. One possibility to bring the first and second period’s profits closer – without changing the effects presented – would be to consider \( z \) to be a mixture of 0 and the uniform component (i.e., a mass point at zero and the rest of the distribution still uniform on \([0, 1]\)).
profit becomes less informative of the employee's ability when \( b \) increases, and the resulting less efficient retention is detrimental to both the firm and, on average (across employee types), to the employee. In particular, as \( b \) increases from zero to infinity, the high type employee's expected surplus (over the outside option \( z \)) from the second period compensation decreases from \( 1/8 \) to \( 1/32 \), while the low type employee's surplus increases from zero to \( 1/32 \), resulting in the average employee surplus decreasing from \( 1/16 \) to \( 1/32 \). Note that the average wage remains the same, but the expected value of it decreases because the value of \( C_2 \) to the employee is

\[
\Pr(C_2 > z) \cdot E(C_2 - z_2 \mid C_2 > z) = E(C_2^2)/2, \text{ i.e., convex in } C_2.
\]

Finally, note that the analysis and all the results of the benchmark case apply whether the employee knows his second period outside option at the beginning of the game or only just before his decision on whether to accept the second period offer. This is because in equilibrium, the employee allocates the effort efficiently in the first period regardless of whether he plans to leave or stay with the firm in the second period.

4.2. Unexpected Measurement Introduced After First Period's Effort Allocation

Now consider a situation in which the measurement is unexpectedly introduced after the first period's effort allocation. This is an off-equilibrium case since we assume that the employee expects to be in the situation of the benchmark case, i.e., he does not expect performance measurement, but the firm then introduces the performance measurement. Thus, the results of this case are going to be used to understand the driving forces and incentives to introduce the performance measurement, and not as predictions in and of themselves. This case is also useful for understanding and predicting what the employees should expect if the firm is unable to commit to whether it would, or would not, introduce a performance measurement mid-game.

In this case, the first period contract, effort allocation, and profits are the same as in the previous case since the firm can do no better and the employees do not expect any measurement to occur. But the second period's contract can now be conditioned not only on \( \pi_t \ (t = 1, 2) \) but also on the measurement realization \( m(x_1) \).\(^{11}\)

\(^{11}\)The case when the second period contract can also be conditioned on the second period's measurement
Given the efficient effort allocation in the first period, which we have since the employee(s) did not expect the performance measurement, Bayes’ rule now gives the following probabilities of the employee being of the high type conditional on the realizations of $\pi_1$ and $m(x_1)$:

\[
\begin{align*}
\Pr(h|\pi_1 = 1 \& m(x_1) = 1) &= 1; \quad \Pr(h|\pi_1 = 1 \& m(x_1) = 0) = \frac{1 + b}{1 + 3b}; \\
\Pr(h|\pi_1 = -b \& m(x_1) = 1) &= 1; \quad \Pr(h|\pi_1 = -b \& m(x_1) = 0) = 0. \quad (12)
\end{align*}
\]

Note that the event $\{\pi_1 = -b \& m(x_1) = 1\}$ does not occur if effort is allocated efficiently, so the Bayes’ rule does not apply in that case. Therefore, $\Pr(h|\pi_1 = -b \& m(x_1) = 1)$ could take any value. We set this value at 1 because a low type employee could never get $m(x_1) = 1$, while a high type employee could get $\pi_1 = -b$ and $m(x_1) = 1$ by not choosing an efficient effort allocation. In other words, this belief assignment is required for an equilibrium to be sequential. However, the results below would also hold if any other value is assumed.

As the new observable in the second period is only $\pi_2$, Equation (7) continues to hold with $P_h$ now depending on both the first period’s profit and on the first period’s measurement of $x$ according to (12) above. Thus, the optimal second period contract offer is $c_21_{\pi_2=1}$, where $c_2 = 1/2$ when $m(x_1) = 1$, $c_2 = 0$ when $\pi_1 = -b \& m(x_1) = 0$ and

\[
c_2 = \tilde{c}_2 \equiv \frac{(1 + b)^3}{2[(1 + b)^3 + 2b^3]} \quad \text{when} \quad \pi_1 = 1 \& m(x_1) = 0. \quad (13)
\]

While the first period decisions and, hence profit, are the same as in the benchmark case, the optimal contract above leads to the expected second period net profit of

\[
E(\pi_2 - C_2) = \frac{1}{8} \frac{(1 + 2b)(1 + b + b^2)}{(1 + b)^3 + 2b^3}. \quad (14)
\]

As $b$ increases from zero to infinity, the above profit decreases from 1/8 to 1/12. Given that the firm now has better information about the employee’s ability in the current case relative to the benchmark, as one would expect, the high type employee retention is higher, the low type realization $m(x_2)$ is considered in Section 5.
employee retention is lower, and the profit and average employee surplus is increased.

4.3. Measurement Introduced Before First Period’s Effort Allocation

Consider now that the measurement was introduced when the first period compensation was set to zero, as in the benchmark case (Section 4.1), but before the employee decides on his effort allocation. Then, the employee knows that the manager will observe the measure \( m(x_1) \) at the end of the first period. Then, it is a dominant strategy for the high type employee to fully distort his first period effort allocation toward the \( x \) component. This is because demonstrating high type increases the expected payoff in the second period in some instances but in no instance decreases it.

Therefore, the expected first period profit is reduced to zero. The expected second period profit increases to \( 1/2 \cdot (1 - 1/2) \cdot 1/2 = 1/8 \), which is the probability of the employee being the high type (given the full effort allocation distortion, such an employee is precisely identified by \( m(x_1) \)) multiplied by the net profit from a retained high type employee multiplied by the probability of retention of a high type employee given the optimal contract. Note that relative to the benchmark case, the expected second period net profit increases due to the better identification and the better retention of the high type employees, but the first period profit suffers. The net effect is that the total net profit of 1/8 becomes even lower than the first period profit of the benchmark case alone, \( E \pi_1 = 1/2 \).

The negative impact of the measurement on the first period effort allocation due to the expected second period contract adjustment leads to the idea that the first period’s incentive to distort effort allocation should be countered in the first period compensation package. We consider this strategy in the following subsection. But first, to illustrate the difficulty of countering the first period distortion, let us derive the high type employee’s benefit of the maximal distortion in the first period relative to no distortion assuming that the firm does not expect a distortion.\(^{12}\)

\(^{12}\)The benefit would be even higher if the firm expects a distortion towards \( x > 1 \), since in that case the high type employee is better identified and, therefore, the second period contract offer following \( m(x_1) = 0 \) would be even lower, while the offer following \( m(x_1) = 1 \) would be the same and equal to \( c_2 = 1/2 \).
The employee's benefit of distortion comes from the possibility of the \( \{ \pi_1 = 1 \& m_1 = 0 \} \) outcome in the first period. In this case, which has probability 1/2 from the high type employee's point of view, the expected second period offer is \( c_2 = \bar{c}_2 \), defined in (13). The employee will accept it with probability \( \bar{c}_2 \), achieving an average surplus, over the outside option and given acceptance, of \( \bar{c}_2/2 \). Thus the expected second period's surplus contribution of this event \( (\pi_1 = 1 \& m(x_1) = 0) \) to the employee's expected utility is \( \bar{c}_2^2/4 \). If the high type employee converts this outcome to \( m(x_1) = 1 \), the surplus is calculated similarly but with \( c_2 \) replaced by 1/2, which results in the expected surplus of 1/16. For example, for \( b = 10 \), the benefit is .0525, which can be seen as quite substantial when compared to the expected second period’s net profit of .07, derived in Section 4.1. This example illustrates that convincing employees not to distort the first period's effort allocation is going to be quite costly to the firm relative to its expected second period's net profit. On the other hand, as we have seen above, not countering the maximal distortion is quite detrimental to the first period profit.

The above analysis was performed under the assumption that the employee does not know the second period's outside option value \( z \) before allocating effort in the first period. Let us now consider what happens if the employee knows \( z \) before allocating effort in the first period. As presented above, we only need to consider the high type employee's effort allocation decisions, as the low type employee can only exert zero effort. In this case, if \( z \geq 1/2 \), the high type employee, in equilibrium, does not value the possible second period offers and thus will, in equilibrium, efficiently allocate his effort in the first period. If \( z < 1/2 \), the high type employee strictly prefers to show that he is of the high type to receive the second period offer with \( c_2 = 1/2 \). Therefore, he will distort his effort maximally. Thus the high type employee distorts his effort with probability of one half, and if he distorts it, he does so maximally. Therefore, the first period profit becomes \( \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} \). Since all employees who can potentially stay distort their effort maximally in the first period, the second period’s net profit is the same as in the case where employees did not know their outside offer, i.e., \( 1/8 \). Thus, the total profit becomes \( 3/8 \), which

\[\text{Note that not everything is the same in the second period: given the employee strategy, the firm optimally updates its belief about the outside option of the employees based on the profit and performance measurement realization. Effectively, } m(x_1) = 0 \text{ signals that the employee expects a high outside option or is of low ability.} \]
is still lower than even the first period profit alone in the benchmark case. Note that in this case, the benefit of distortion is not the same for all employees who distort their effort allocation: those with $z$ close to $1/2$ are almost indifferent between distorting and not distorting their effort allocation, while those with $z$ below $c_2$ offered in the $\{\pi_1 = 1 & m(x_1) = 0\}$ outcome have the highest incentive to distort.

4.4. Measurement Introduced Before First Period’s Contract

We now turn to the main case of the contract design when all the parties know at the beginning of the game that the measurement was introduced. Since distorting $x$ upward from the efficient allocation in the first period reduces the first period’s expected profits due to the increased $x_1$, the principal may try, at least partially, to counteract this first period’s distortion in the first period’s contract by a combination of (a) paying for $\pi_1 = 1$ when $m(x_1) = 1$, (b) paying for $\pi_1 = 1$ when $m(x_1) = 0$, or (c) paying for $m(x_1) = 0$ when $\pi_1 = -b$. Option (c) is clearly worse than (b) as it allocates more spending toward low type employees and provides less incentive to efficiently allocate effort to increase profit. The relative optimality of the first two instruments is less straightforward. Still, for $b > 4$, an increase in the weight on outcome in (a) is counter-productive (increasing $x$ increases the probability of this outcome when $b \geq 4$). Therefore, for large $b$, the optimal contract will involve a positive pay only on the outcome in (b). This is intuitive as such a contract gives employees the incentive to move $x$ down towards the efficient amount both through conditioning on $\pi_1$ being high (profit sharing incentivizes an efficient allocation) and through conditioning on $m(x_1)$ being low (an incentive to reduce $x$ from whatever level it would be otherwise set).

Consider now the conditions under which a high type employee prefers no distortion to the maximal distortion given the first period incentive $C_1 = c1_{\pi_1=1,m(x_1)=0}$ assuming that the

But this feature does not affect the expected profit. This is because the probability of a high type given this realization is less than $1/2$ (all low-ability but not all high-ability employees have this measurement). Therefore, the firm will not be willing to offer more than $c_2 = 1/2$. This means that the high type employees who planned to leave will still leave and, therefore, it is still optimal for the firm to offer precisely zero compensation having observed $m(x_1) = 0$.

$^{14}$For a high type employee the probability of the outcome $\{\pi_1 = 1 & m(x_1) = 1\}$ is $2x_1\frac{b+x_1(2-x_1)}{1+b}$, which is increasing in $x_1 \in [0,2]$ when $b \geq 4$. 

19
firm expects no distortion. Note that the maximal distortion redistributes the event \( \{ \pi_1 = 1 \& m(x_1) = 0 \} \), which for a high type employee with no distortion has probability 1/2, to the event \( \{ \pi_1 = 1 \& m(x_1) = 1 \} \) with (total additional) probability \( \frac{b+\sqrt{2}-1}{2(1+b)} \) and to the event \( \{ \pi_1 = 0 \& m(x_1) = 1 \} \) with (total additional) probability \( \frac{2-\sqrt{2}}{2(1+b)} \). Each of these cases leads to an increase in the second period’s offer from \( c_2 = \tilde{c}_2 \) derived in Subsection 4.2 to 1/2, and results in a loss of the first period’s incentive \( c \). Note that the benefit of increased pay in the second period needs to be counted net of foregoing the outside option and taking into account the probability of staying with the firm. In other words, the expected benefit of a contract with the expected pay of \( w \) is \( \int_0^w (w - z) \, dz = w^2/2 \). Thus, to prevent a high type employee deviation to the full distortion of \( x \), we need

\[
c \geq (1/2)^2/2 - \tilde{c}_2^2/2 \equiv \frac{1}{8} - \tilde{c}_2^2/2.
\]  

(15)

Define

\[
\tilde{c} \equiv \frac{1}{8} - \frac{\tilde{c}_2^2}{2}.
\]  

(16)

Let us now consider the possibility of the employee partially distorting his effort allocation. It turns out that while given the incentive \( \tilde{c}_{1\pi_1=1,m(x_1)=0} \), a high type employee chooses no distortion, any smaller incentive leads to the maximal distortion. This is because the high type employee’s expected payoff as a function of the first period’s \( x_1 \) is tilde-shaped: the employee’s payoff as a function of \( x_1 \) has a maximum at \( x_1 \) below \( 1/4 \) for incentives strictly higher than \( \tilde{c} \). For incentives lower than \( \tilde{c} \) the global maximum of his payoff is \( x_1 = 2 \). The Appendix has the detailed proof that preventing maximal distortion prevents any distortion towards \( x > 1 \). Again, note that if the firm expects some deviation, then the employee has an even higher incentive to distort, because \( c_2 \) offered in the event \( \{ \pi_1 = 1 \& m(x_1) = 0 \} \) becomes lower. Therefore, (16) establishes a lower bound on the first period’s contract incentive necessary to prevent the high type employees from maximal distortion of their effort allocation in the first period toward the \( x \) component.

Consider now if it is beneficial for the principal to incentivize no distortion in the first period. (As we established above, this needs to be compared to the maximal distortion.) Given no effort distortion in the first period, the low type employee will get \( \tilde{c} \) with probability \( \frac{b}{1+b} \), and the high
type employee will get $\tilde{c}$ with probability 1/2. The expected cost for the principal of incentivizing no distortion in the first period is therefore $\frac{1+3b}{4(1+b)}\tilde{c}$. The expected gross benefit of incentivizing no distortion (relative to full distortion) in the first period is 1/2.

The employee’s effort allocation distortion in the first period would help the profit in the second period due to allowing a superior second period’s contract (due to the employee’s ability being fully revealed). The value of this to the principal is $\frac{1}{2} - \frac{1}{8} \frac{(1+2b)(1+b+b^2)}{(1+3b+3b^2+3b^3)}$ (the first term is the second period profit given maximal first period distortion derived in Section 4.3; the second one is the second period profit given no first period distortion derived in Section 4.2). We can then obtain that the benefit of countering the maximal distortion is higher than the total cost of doing so. Given that the principal counters the distortion, the first period’s profit (net of the cost of preventing the distortion) becomes $\frac{1}{2} - \frac{1+3b}{4(1+b)}\tilde{c}$ and the total net profit becomes

$$\frac{1}{2} - \frac{1+3b}{4(1+b)}\tilde{c} + \frac{(1+2b)(1+b+b^2)}{8(1+3b+3b^2+3b^3)}, \quad (17)$$

which is smaller than the total net profit in the benchmark case. For example, for $b \to \infty$, the net profit converges to 1/2, which is smaller than the net profit when the measurement was not possible (which tends to 9/16), but greater than the net profit would be if the firm did not counter the first period distortion ($0 + 1/8 = 1/8$). Summarizing the results of this subsection, we obtain the following proposition:

**Proposition 1:** Suppose that $b$ is sufficiently large (e.g., $b > 4$), and that employees do not know their second period outside option $z$ before allocating their first period effort. Then, if the measurement of one of the effort components is implemented at the start of the game, we have:

1. The firm chooses to incentivize employees not to distort their first period effort allocation. This results in a loss in the first period’s profit but with no change in efficiency.

2. The firm uses the first period’s measurement for a better retention of high type employees in the second period. This is (on average) beneficial to the high type employees and detrimental to the low type employees, and results in increased second period’s profits.
3. The net result of the two effects above is that the total net profit decreases, while the employees, on average, are better off.

4. Social welfare increases.

The proposition states the results for $b$ large, but we could not find values of $b > 0$ such that any of statements of this proposition did not hold. The proposition states that there is no loss of efficiency in the first period and that social welfare increases.

We now discuss now generality of these two results, when productive employees are heterogeneous in the first period. As we show below, the incentive to distort may be too large and the firm may then choose not to counter it. This can then lead to the possibility of welfare losses due to the lower productivity. In fact, if the productive employees are heterogeneous in the first period, there could be no way to design an incentive to not distort effort. To show this within this model, let us get back to the possibility that the employees have some information about their second period’s outside option before they choose their first period’s effort allocation.

In contrast to the previous model analysis, now assume that at the beginning of the first period, the employees know $z$ precisely. In this case, if measurement is introduced, employees who know that they are going to leave for sure, e.g., those with $z > 1/2$, do not have an incentive to distort $x_1$ unless the first period contract provides them with an incentive to distort effort.

Therefore, since potentially half of the productive employees will leave, when designing the incentive structure to induce no (or less) distortion by the other employees, the firm needs to make sure that the incentive is not strong enough to fully distort the effort allocation of the employees who will leave. The firm can use a combination of a positive weight on the first period profit $\pi_1$ to incentivize no distortion of those planning to leave, and an additional incentive through a positive weight on $m(x_1) = 0$.

The optimal wage for $\pi_1$ cannot be higher than 2, as that is the maximum possible output over two periods of a high ability employee. Therefore, the maximal incentive that it provides tends to zero as $b$ tends to infinity. This implies that if any weight is put on $m(x_1) = 0$ in the first period contract, as $b$ tends to infinity, the employees that are expecting to leave the firm for
sure (i.e., those with $z > 1/2$) will distort their effort allocation maximally toward the lowest $x_1$ (i.e., towards $x_1 = 0$). As the firm benefits from identifying the type of the employees who are staying with the organization for better retention efficiency in the second period, it prefers those employees who will stay to distort their effort allocation as opposed to the employees who will leave. Therefore, not to have the employees who will leave distort their effort allocation fully, the optimal weight on $m(x_1) = 0$ must tend to zero as $b$ tends to infinity. That is, as $b$ tends to infinity, the first period contract approaches the form $p1_{x_1 = 1}$ (for some non-negative $p$).

Again, as the optimal $p$ is bounded from above (as it is by 2), the incentive it provides not to distort the effort allocation tends to zero as $b$ tends to infinity. Therefore, as $b$ tends to infinity, almost all employees who will stay distort their effort allocation maximally. (Only those with $z$ close to 1/2 do not care much about the second period offer and therefore do not distort their effort given a small incentive.) It then follows that as $b \to \infty$, the first period profit converges to 1/4 and the second period profit converges to 1/8, and we have the following proposition (see Appendix for a more formal proof):

**Proposition 2:** Suppose $b \to \infty$ and that employees know their second period outside option $z$ before allocating effort in the first period. Then, if the measurement is introduced,

1. The mass of employees who end up staying in the second period and distort their effort in the first period maximally toward the measured component converges to 1/2.
2. The mass of employees who leave and distort their first period effort in an amount converging to zero converges to 1/2.
3. The total net profit and the social welfare are lower than if the measurement were not introduced.

To see the social welfare result, note that without the measurement of $x$ the total production created by the organization is the share of the high type employees being employed times their productivity, which is $\frac{1}{2}$ in the first period and $\frac{z_2}{2}$ in the second period for a total of $\frac{1+z_2}{2}$. Adding
to this the expected total outside option obtained by the employees who leave, one obtains the total production created as

$$
\frac{1 + \hat{c}_2}{2} + \frac{1 - \hat{c}_2^2}{4} + \frac{1}{4} \cdot \frac{1}{1 + b} + \frac{1}{4} \cdot \frac{b}{1 + b} \left[ 1 - \left( \frac{b + \hat{c}_2}{1 + b} \right) \right],
$$

(18)

where the second term represents the expected value of the outside option of the high type employees who leave in the second period, the third term represents the expected value of the outside option of the low type employees who did not get positive output in the first period (all of these employees leave in the second period), and the fourth term represents the expected value of the outside option of the low type employees who received positive output in the first period.

When $b \to \infty$, we have $\hat{c}_2 \to 1/4$ and, therefore, the total production tends to $35/32$. With measurement, when $b \to \infty$, half of the high type employees maximally distort their effort allocation in the first period, but all of these stay in the second period, which results in the total production of the firm being $1/2(1/2 + 1/2) = 1/2$. Adding to this the expected total outside option obtained by the employees who leave, one obtains the total production created converging to

$$
\frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{3}{4} + \frac{1}{2} \cdot \frac{1}{2} = \frac{15}{16},
$$

(19)

where the second term captures the expected value of the outside option of the high type employees who leave in the second period, and the third term is the expected value of the outside option of the low type employees in the second period. (All leave in the second period as the measurement reveals that they are low type.)

Comparing (18) for $b \to \infty$ with (19), we obtain that the social welfare with measurement is lower than without measurement.

Note that in a more complete model, which would account for the origins of the outside option and the benefit to other firms of employees leaving this firm, one can argue that any information asymmetry between firms and employees is welfare reducing, as it results in inefficient allocation of employees across firms. (Essentially, wages may be only reallocating surplus from firms to employees, but unequal wages, due to different information that different firms have about the
employee, means that the employee may stay at a less preferable firm.) Such consideration of the social desirability of retention efforts is beyond the scope of this paper.

Proposition 2 shows how some heterogeneity among high type employees (namely, their knowledge of different outside option values in the second period) results in the inability of the firm to prevent employee effort distortion in the first period by up to one half of the employees. The only reason most employees did not distort effort is that half of the employees were essentially homogenous: the effort allocation of those who know they will be leaving for sure is not affected by the exact value of the outside option. With more heterogeneity in the employee beliefs about the probability of leaving, one could obtain that, although in aggregate the firm may incentivize any proportion of $x$ and $y$ efforts, it may be that nearly all employees maximally distort their effort. The Appendix provides an example of such a model variation.

Although in all the different set-ups above, we have that measurement, on average, benefits the high type employees, one can now see that introducing a first period outside option or competition between organizations, so that the employees may share the expected welfare surplus from employment through the first period offer, could lead to the introduction of measurement decreasing the payoffs of the employees and the organization alike.

5. Measurement of Effort in Both Periods

In this section we investigate what happens when the measurement of component $x$ can be done in both periods. In other words, we consider the analysis of the previous section with the added possibility of $m(x_2)$ being also available. As we will see, this brings additional complexity into the analysis without affecting the main messages of the previous section. For analytical tractability, we perform this analysis focusing on large $b$.

5.1. Unexpected Measurement Introduced After First Period’s Effort Allocation

Consider first the case in which there is an unexpected introduction of the measurement after the first period’s effort allocation. Just as in the analysis of Section 4.2, in the second period,
uncertainty about the employee’s type remains only if \( m(x_1) = 0 \).

In this case, if \( m(x_2) \) is also observable and absent concerns about the second period effort allocation, the principal may want to put a positive weight on the second period measurement for a better retention of the high type employees. This is so when \( m(x_2) \) is informative of the employee’s ability on top of the information derived from the first period observables (\( \pi_1 \) and \( m(x_1) \)) and the second period profit realization \( \pi_2 \). If effort is efficiently allocated in the first period, this would be when \( \pi_1 = 1 \) and \( m(x_1) = 0 \) (otherwise, first period observations reveal the type). Thus, one can expect a strictly positive weight on \( 1_{m(x_2)=1} \) in the second period contract when (and only when) \( m(x_1) = 0 \).\(^{15}\) The only reason that this weight may not be positive is if it could lead to a second period’s effort allocation distortion and the manager finds this suboptimal. The outcome of this trade-off depends on the probability of the employee being a high type: if this probability is lower, the benefit of identifying the high type (and not spending on the low type) is higher.

Similarly to the proof of Proposition 2, for any positive incentive on \( m(x_2) \), as \( b \to \infty \), a high type employee will distort the effort allocation in the second period maximally. Since maximal distortion leads to zero productivity, for large \( b \) the weight on \( m(x_2) \) must be small and, therefore, most of the pay comes from the weight on \( \pi_2 \) (clearly, a weight on \( \pi_2 \) is better than a constant component, so the constant will not be used). Thus, asymptotically, the second period outcomes are as if the measurement did not exist in the second period. That is, the results of Section 4.2 apply here asymptotically for large \( b \).

For an illustration, consider the objective function of a high type employee who has a contract \( c_2 1_{\pi_2=1}(1 + w_2 m(x_2)) \) for some \( c_2 > 0 \) and \( w_2 > 0 \). If the employee chooses effort allocation \( x_2 \), his expected wage is:

\[
\hat{w}_{2e} = \frac{(b + x_2(2 - x_2))(1 + w_2 x_2^2) c_2}{1 + b}.
\]

\(^{15}\) Note that this could look as a special treatment for apparent (i.e., as measured) substandard employees to "encourage improvement." This is not exactly the case because achieving \( m_1(x_1) = 1 \) leads to not lower expected compensation in the second period for retention reasons.
The first order condition with respect to $x_2$ implies that the optimal $x_2$ is

\[
x_2 = \min \left\{ 2, \frac{2w_2 - 2 + \sqrt{(2w_2 + 1)^2 + 3(1 + w_2^2b)}}{3w_2} \right\}.
\]

(21)

Note that, asymptotically, for small $w_2$, the optimal $x_2$ equals $1 + \frac{1+b}{4}w_2$. Using the decision rule given by (21), we can now write the first-order conditions with respect to $c_2$ and $w_2$ on the second period profit in the case \{\pi_1 = 1 \& m(x_2) = 0\} using the probability of the high type given in (12). The analysis is presented in the Appendix. For $b = 10$ we can obtain $c_2 = .20$ and $w_2 = .02$, resulting in $x_2 = 1.05$.

5.2. Unexpected Measurement Introduced After First Period Contract but Before First Period Effort Allocation

In this case, exactly as in the case of Section 4.3, all employees maximally distort their effort allocation in the first period. The first period’s measurement $m(x_1)$ is then fully informative about the employee’s type and, thus, exactly the same outcomes follow. Note that in the first period, employees may expect the second period offer following $m(x_1) = 0$ to have some positive weight on $m(x_2)$, but since any significant weight on $m(x_2)$ would result in the maximal distortion (for sufficiently large $b$) and therefore, zero expected profits, the second period offer that the employees may expect following $m(x_1) = 0$ is strictly worse than the second period offer that they expect following $m(x_1) = 1$, which is $1/2 \cdot \pi_2=1$. Therefore, the maximal effort distortion in the first period guaranteeing $m(x_1) = 1$ is strictly optimal for the employees. Again, note that the first period profit loss due to this distortion in the first period is $1/2$.

5.3. Measurement Introduced Before the First Period Contract

In this case, it follows from the results in the previous section that the firm will want to prevent maximal distortion. As the second period offer given no distortion is asymptotically the same as if the measurement in the second period did not exist, the results of Section 4.4 hold here asymptotically. We thus confirm all the results of Propositions 1 and 2, except that the firm uses
the second period measurement for retention in a small amount.

6. Discussion

One may find it curious to reflect on the Finnish education achievement puzzle from the point of view offered by the above model. As demonstrated by eventual student abilities, Finland was able to improve performance by largely eliminating both the teacher and student performance measurements (see Darling-Hammond 2010).

A tenure system may also be considered as a commitment to limit the use of measurements. For example, a standard justification for life-time appointments of justices is that otherwise they could be swayed by unscrupulous decision makers. Our research puts it in a different light: even if the supervisors are fully benevolent (they only have the objective of maximizing public welfare) and are fully rational, they may not be able to not interfere and not distort the socially efficient decisions of the employees.

The considerations illustrated by the model easily apply to not-for-profit organizations. In such organizations, the productivity/output \((a)\) in fact may not be observable at all in the absence of measurement, yet the organization cares about it by definition. Note that any measurement of the efforts or productivity components might have the property of not being completely unbiased between the different components of output, i.e., almost always under or over-weigh one type of the input efforts. Furthermore, both the employees and the management of such organizations may have pride in their work, which leads to the employees preferring to optimize \(a\) absent other incentives (in the model, due to a small weight on the total output of the firm). The management, while by definition may be interested in the “good deeds” \(a\) brought to the society, also prides itself in the amount of work done by its particular organization, and therefore values the retention of the productive (i.e., high-\(a\)) employees.

Note further that retention by itself is not necessarily socially desirable. In the model we formulated, it is efficient (socially desirable) to retain an employee if and only if the employee’s outside option \(z\) is less than the productive output \(a\) he can generate. Therefore, it is efficient
to offer a better contract (higher base pay) to those employees who are expected to be more productive. As we have shown, the firm may be worse off, although the employees and the social welfare are better off in our model when the principal is willing and able to incentivize employees not to distort their effort allocation in the first period (although social welfare is lower in the model variation where employees observe their second period’s outside offer before the first period). However, another interpretation is that \( z \) comes from the idiosyncratic employee preferences for outside attributes such as job location plus a job offer from an organization with exactly the same production function. In this case, given equal pay (contracts), employees would produce the same \( a \)'s but realize the best possible location choice, i.e., the location-inconvenience cost would be minimized. Employee retention then makes utility generated by locations inefficiently distributed while only shifting \( a \)'s between organizations. Then, the management’s work (introducing measurements and fine-tuning contracts) could make both the organization and, on average, the employees worse off (although some employees may be better off). The model can then be the most strikingly characterized as exploring the tug of war between the employees’ pride in their work being a force toward the efficient society and the managerial pride in their work being a force against efficiency and toward a distortion of both the employees’ effort allocation and the employee locations.

Of course, in many other situations, the problem of effort inducement and employee retention is the main problem an organization faces, and the problem of optimal effort allocation between unobserved (not well-measured) components is not as essential. In those cases, performance measurements could benefit the organization and an average employee. The point of this paper is not that performance measurement is always or even usually counter-productive, but that it could be, hence one should consider the implications of measurement before its introduction. In other words, one should have an idea of how data is to be used before collecting it.
Appendix

Sub-Optimality of Partial First Period Distortion:

Consider the high type agent’s first period effort allocation decision $x_1$ given the first period contract $c1_{π_1=1&m(x_1)=0}$ and the expected second period offer $\frac{1}{2}1_{m(x_1)=1} + \bar{c}_21_{π_1=1&m(x_1)=0}$. The employee’s value of increasing $x$ above $x = 1$ is coming from the increased probability of $m(x_1) = 1$ and the associated second period equilibrium pay increase from $\bar{c}_2$ to $1/2$:

$$\text{Benefit} = \left(\frac{x}{2} - \frac{1}{2}\right) \left(\frac{(1/2)^2}{2} - \bar{c}_2^2\right).$$

This benefit comes at a cost of lower first period pay

$$\text{Cost}_{\text{first period}} = \left(\frac{1}{2} - b + a \left(1 - \frac{x}{2}\right)\right) c,$$

where $a = x(2 - x)$,

and of the possibility that the wage is reduced from $\bar{c}_2$ to 0 (event $\{π_1 = -b & m(x_1) = 0\}$ becomes possible), which is valued by the employee at

$$\text{Cost}_{\text{second period}} = \frac{1 - a}{1 + b} (1 - x/2)\bar{c}_2^2/2.$$

Subtracting the two costs from the benefit, we obtain the employee’s objective function as

$$f(x, c) = \frac{1}{16} \frac{(x - 1)(1 + b + 4(x^2 - 3x - b + 1)(\bar{c}_2^2 + 2c))}{1 + b}.$$

By definition, this $f(x, c) = 0$ at $x = 1$. At $c = \bar{c} \equiv \frac{1}{8} - \frac{\bar{c}_2}{2}$, it simplifies to

$$f(x, \bar{c}) = -\frac{(x - 1)^2(2 - x)}{16(1 + b)},$$

which clearly achieves the maximum (of 0) at $x = 1$ and $x = 2$. Thus, the incentive that is just enough to prevent the employee from preferring no distortion to maximal distortion, makes the
employee strictly prefer no distortion to any intermediate distortion. Furthermore,

\[ \frac{\partial f(x,c)}{\partial c} < 0 \quad \text{and} \quad \frac{\partial^2 f(x,c)}{\partial c \partial x} < 0 \quad \text{for} \quad x \in (1, 2]. \tag{27} \]

In other words, \( f(x,c) \) decreases in \( c \) and the speed (by absolute value) of this decrease increases in \( x \). This implies that if for some \( x_1 < x_2 \), we have \( f(x_1,c_1) \leq f(x_2,c_1) \), then for any \( c < c_1 \), we have \( f(x_1,c) < f(x_2,c) \). Applying this to \( c_1 = \bar{c} \) and \( x_2 = 2 \), we obtain that for any smaller incentive than \( \bar{c} \) (i.e., for \( c < \bar{c} \)), we have that the effort allocation \( x = 2 \) is preferable to any effort allocation \( x \in [1, 2) \). Conversely, any incentive larger than \( \bar{c} \) results in \( x < 1 \).

**Proof of Proposition 2:**

Fix arbitrary \( p_m > 0 \) and \( w > 0 \), and consider a high type employee decision on his first period effort allocation given the first period contract \( p_1 \pi_1 = 1 - w m(x_1) \) with \( p \leq p_m \). We first prove that for sufficiently high \( b \), an employee with \( z > 1/2 \) (i.e., an employee who is sure to leave in the second period) will distort his first period effort allocation to \( x_1 = 0 \). To see this, note that

\[ \frac{d}{dx_1} \text{Prob}(\pi_1 = 1) = \frac{2 - 2x}{1 + b} < \frac{2}{(1 + b)^2}, \tag{28} \]

which means that for \( b > \frac{4p_m - w}{w} \), the above derivative becomes smaller than \( \frac{w}{2p_m} \). Therefore, for \( b > \frac{4p_m - w}{w} \), the expected first period pay of a high type employee decreases in \( x_1 \) for all \( x_1 \in (0, 2) \) given the contract considered above. Therefore, given the above first period contract, an employee who knows for sure he will leave (i.e., who does not care about the second period contract) will choose \( x_1 = 0 \) if \( b > \frac{4p_m - w}{w} \).

Facing the above choice of high type employees with \( z > 1/2 \), the firm would prefer to instead have employees with \( z > 1/2 \) not distort their effort allocation from the optimal one \( (x_1 = 1) \) at all, while having the rest of the employees distort the effort allocation maximally upward (i.e., choose \( x_1 = 2 \)). Since the latter outcome is achieved with the zero-pay contract, the former contract cannot be optimal. Thus \( w \) of the optimal contract must tend to zero as \( b \) tends to infinity. (Otherwise, its optimality for sufficiently large \( b \) is contradicted by the above observation.
that the zero pay contract is strictly better.)

Now, fix arbitrary \( z_{\text{max}} < 1/2 \) and \( x_1^* \in [0, 2) \), and consider a high type employee with the second period outside option \( z \leq z_{\text{max}} \). As derived in the main text, the second period contract provides the high type employee the expected second period pay of \( 1/2 \) following \( m(x_1) = 1 \) and strictly lower following \( \{m(x_1) = 0 \ \& \ \pi_1 = -b\} \) or \( \{m(x_1) = 0 \ \& \ \pi_1 = 1\} \). Let \( c_m < 1/2 \) be the highest pay out of the latter two. Then increasing \( x_1 \) from \( x_1^* \) to 2 increases the expected pay (conditional on the employee accepting the offer) in the second period by at least \( (1/2 - c_m) \) with probability \( (1 - x_1^*/2) \) and never decreases it (once \( x_1 = 2 \), the expected pay is always \( 1/2 \)). The employee values this expected increase in the second period offer by at least \( (1 - x_1^*/2) \min\{1/2 - z_{\text{max}}, 1/2 - c_m\} \). Remember that we have established above that the optimal \( w \) of the optimal first period contract \( p_1 \pi_1 = 1 - w m(x_1) \) tends to zero as \( b \) tends to infinity, and that \( p \leq 2 \) in an optimal contract. Therefore, the marginal penalty from increasing \( x_1 \) coming from the first period’s optimal contract tends to zero as \( b \to \infty \). Therefore, for large enough \( b \), all high type employees with \( z < z_{\text{max}} \) will choose \( x_1 = 2 \).

We have thus established the first part of the proposition. Note also that it then follows that as \( b \to \infty \), almost all the employees with \( z < 1/2 \) stay in the second period. (To be precise, for all \( z_{\text{max}} < 1/2 \), for sufficiently high \( b \), all employees with \( z \leq z_{\text{max}} \) will maximally distort the first period effort allocation and will be guaranteed the expected pay of \( 1/2 \) in the second period, and therefore, stay for sure.) Thus, employees with \( z > 1/2 \) form “almost all employees who leave.” Effort distortion of employees with \( z > 1/2 \) has to tend to zero as \( b \to \infty \), because since nearly all the rest of the employees distort maximally, zero-pay contract (and the resulting no-effort distortion by the employees with \( z > 1/2 \)) would be better otherwise. This establishes the second part of the proposition. The last part and the social welfare result have been calculated in the main text (based on the first two). This concludes the proof of the proposition.

**CASE WHEN EMPLOYEES HAVE PARTIAL INFORMATION ABOUT SECOND PERIOD OUTSIDE OPTION:**

Assume now the following information structure the employee has at the beginning of the game about his second period outside option. Assume at the beginning of the game that the
employee has a signal \((k, \sigma)\) about his second period outside option, where \(k \in \{0, \ldots, K\}\) represents the reliability of \(\sigma\) and is uniformly distributed on \(\{0, \ldots, K\}\), while \(\sigma\) is equal to \(1_{z>1/2}\) with probability \(k/K\) and is a random draw from \(\{0, 1\}\) otherwise. Each of these \(2K + 2\) possible (and equally likely) signals corresponds to a different value that the high type employee receiving such a signal places on the second period equilibrium offer.

Let \(c_2^*\) be the equilibrium second period \(c_2\) when \(\pi_1 = 1 \& m(x_1) = 0\) and the manager expects efficient effort allocation in the first period (by the same argument as in Section 4.2, if the manager expects a distortion, \(c_2\) will be lower). As before, by maximally distorting the \(x\) component of effort upwards, the high type employee can increase \(c_2\) and, therefore, the value of the second period offer (since it can ensure \(\pi_2 = 1\) by allocating effort efficiently in the second period), to \(1/2\). Again, as before, the value of this action equals the probability of accepting the contract with \(c_2 = 1/2\) times the expected increase in the payoff (which is \(1/2 - \max\{z, c^*_2\}\)) conditional on accepting the second period offer. The difference lies in the evaluation of the probability of accepting the offer, which now is a function of the signal \((k, \sigma)\).

Signal \((k, \sigma)\) is uninformativ e with probability \(1 - k/K\). In this case, \(z\) is uniformly distributed on \((0, 1)\) and therefore, the value of increasing \(c_2\) from \(c_2^*\) to \(1/2\) is \(v^* \equiv (1/2 - c_2^*)c_2^* + \int_{c_2^*}^{1/2} (1/2 - z) \, dz\). With probability \(k/K\), the signal is informative. In this case, if \(\sigma = 1\), the employee is sure to leave even with \(c_2 = 1/2\) and, therefore, has no value of increasing the second period offer. However, if \(\sigma = 0\) (and the signal is true), the employee is sure to stay and values increasing the offer at \(2v^*\). Aggregating the above (taking into account the probability of the event \(\{\pi_1 = 1 \& m(x_1) = 0\}\)), we find the values \(v_{k, \sigma}\) the high type employee with signal \((k, \sigma)\) places on increasing the second period offer. These values are ordered as:

\[
v_{K, 1} < v_{K-1, 1} < \cdots < v_{0, 1} < v_{0, 0} < \cdots v_{K, 0}.
\]

These values represent the incentive the employee has to maximally distort his effort allocation toward \(x = 2\) given signal \((k, \sigma)\). Let \(dv\) be the smallest distance between any pair of the above values. Again, if the manager believes the effort allocation is distorted, \(dv\) will be greater. Thus,
$dv > 0$ represents the lower bound on how different incentives are between the high type agents with different signals.

Similar to the argument leading to Proposition 2, one can now prove that if $b$ is large enough, a potentially optimal first period contract (e.g., one where the weight on $\pi_1$ is bounded from above by a number independent of $b$) may only incentivize no-effort distortion by agents with one of the signal possibilities. The rest will be distorting $x$ either maximally upward or maximally downward. Thus, for large $K$, almost all high type agents distort their first period effort allocation maximally, and we have the following proposition.

**Proposition 3:** If employees have some information about their second period outside option, it is impossible to profitably prevent maximal effort distortion by nearly all employees in the first period.

**Unexpected Measurement of the Second Period Effort After the First Period’s Effort Allocation:**

The optimal $x_2$ maximizes (20). Using implicit differentiation of the first order condition $4 + 4w_2x_2 - 4x_2 - 3w_2x_2^2 + w_2b = 0$ on the optimal $x_2$, we obtain

$$\frac{\partial x_2}{\partial w_2} = \frac{4x_2 - 3x_2^2 + b}{4 - 4w_2 + 6w_2x_2}.$$  (30)

The firm’s expected profit in the second period given $\pi_1 = 1&m(x_1) = 0$ is

$$E(\pi_2|\pi_1 = 1&m(x_1) = 0) = \phi_{10}\frac{a + b}{1 + b}c_2(1+x_2w_2/2)(1 - \frac{a + b}{1 + b}c_2(1+x_2w_2/2)) + (1 - \phi_{10})\frac{b}{1 + b}c_2^2[0 - \frac{b}{1 + b}c_2],$$  (31)

where $\phi_{10}$ is the probability of high type conditional on $\pi_1 = 1&m(x_1) = 0$ which is $\phi_{10} = \frac{1 + b}{1 + 3b}$ per (12), $a = x_2(2 - x_2)$, and $x_2$ is a function of $w_2$, which is implicitly defined by the FOC on the employee’s objective function (or explicitly by (21)).

The optimal $c_2$ and $w_2$ can then be found by taking the derivatives of (31) with respect to $c_2$ and $w_2$, using (30) for the derivative of $x_2$ with respect to $w_2$, and making those derivatives
equal to zero.
References


