

NBER WORKING PAPER SERIES

SUBJECTIVE EVALUATION OF PERFORMANCE THROUGH INDIVIDUAL EVALUATION INTERVIEW:
EMPIRICAL EVIDENCE FROM FRANCE

Marc-Arthur Diaye
Nathalie Greenan
Michal Urdanivia

Working Paper 12979
<http://www.nber.org/papers/w12979>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2007

This paper was presented at the Conference on Analysis of Firms and Employees (CAFE 2006) held September 29-30, 2006 in Nuremberg, Germany. We gratefully acknowledge the financial support provided by the Institute for Employment Research (IAB), the Data Access Center (FDZ-BA/IAB), The Deutsche Forschungsgemeinschaft (German Research Foundation), their Research Network "Flexibility in Heterogeneous Labour Markets", the Alfred P. Sloan Foundation and the National Science Foundation. We are deeply indebted to Emmanuel Duguet and to Lynn Zucker who help us to improve this paper. We thank the participants at the CAFE 2006, at the BETA-Cereq (University Louis Pasteur) seminar on labour Market, at the PSE-Jourdan seminar on labour market and inequalities (especially Andrew Clark) and at the TEAM Seminar of the Centre d'Economie de la Sorbonne for their comments. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

© 2007 by Marc-Arthur Diaye, Nathalie Greenan, and Michal Urdanivia. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Subjective Evaluation of Performance Through Individual Evaluation Interview: Empirical evidence from France

Marc-Arthur Diaye, Nathalie Greenan, and Michal Urdanivia

NBER Working Paper No. 12979

March 2007

JEL No. M5,M50,M54,M55

ABSTRACT

Individual evaluation interviews have become a widespread practice. 52% of employees in French manufacturing firms over 50 employees declared an annual individual evaluation interview in 1997. However whereas the problem of constructing an optimal contract with subjective evaluation (which is defined simply as a signal in most papers) receives a large attention, firm-level evaluation interviews are strikingly left aside from economic analysis. This paper aims at identifying the underlying logics of individual evaluation interviews in the case of individual production and of team production. Especially, it aims at analyzing the relationships between effort, wage distribution within the firms and individual evaluation interviews. From a theoretical standpoint, three papers by Alchian and Demsetz (1972), by Che and Yoo (2001) and by MacLeod (2003) are closely related to our paper and from an empirical point of view, a paper by Engellandt and Riphahn (2004). Our theoretical analysis allows to derive testable predictions regarding the effect of individual evaluation interviews on productive and cognitive effort, on work overload and on wage setting. Using a matched employer / employee survey on computerisation and organisational change (COI), we are able to test part of these predictions and to corroborate them. First, evaluation interviews have a positive impact on productive and cognitive effort. Second, evaluation interviews increase effort through two effects: the classical incentive effect and also a selection effect. Third, the selection effect is stronger in the case of individual production compared with the case of team production. Fourth, evaluated employees earn more than employees in a classical incentive scheme and fifth, evaluated workers have a better knowledge of the rules driving wage setting.

Marc-Arthur Diaye
29 Promenade Michel Simon
93166 Noisy-le-Grand Cedex
France
marc-arthur.diaye@cee.enpc.fr

Michal Urdanivia
Maison des Sciences Economiques (CERMSEM)
106-112 boulevard de l'Hopital
75013 Paris, France
michal.w-urdanivia@univ-paris1.fr

Nathalie Greenan
29 Promenade Michel Simon
93166 Noisy-le-Grand Cedex
France
nathalie.greenan@cee.enpc.fr

"If the economic organization meters poorly, with rewards and production only loosely correlated, then productivity will be smaller; but if the economic organization meters well productivity will be greater."

Alchian and Demsetz (1972, page 779)

1. Introduction

A main contribution of linked longitudinal employer / employee data is to provide a decomposition of wage rates into components due to individual heterogeneity and to firm heterogeneity. In France, Abowd, Creecy and Kramarz (2002) show that the person effect and firm effect account respectively for 70% and 20% of the variation of wages. The person effect component is bigger in France than in the US where it represents half of the wage variation.

This indicates that the devices used by firms to attract and / or select workers with specific characteristics play a central role in determining the firm's wage structure. However, these devices have not been investigated thoroughly by economic analysis. In this paper, we are going to assess individual evaluation interview, an HRM practice that could contribute to the two goals of selecting workers and stimulating their effort. In France, 52% of employees with more than one year of seniority in manufacturing had been evaluated at least once in 1997. At that time, evaluation interviews were not regulated at the national or at the industry level.

As an HRM practice, the function of evaluation interviews is not clear cut. Sometimes viewed as formal performance appraisal systems, evaluation interviews often use complex evaluation grids referring to loosely defined behavioural characteristics as well as to precisely defined goals and measured criteria.

To assess evaluation interviews, it is useful to analyse them theoretically and to investigate empirically how they are implemented within firms.

Intuitively, individual evaluation interviews are used to assess performance once the employee has undertaken her or his task. We will use here the term of *ex post* evaluation interviews. But, if we refer to a classic wage setting mechanism, there is no need for evaluation. The incentive wages drives the employee towards the level of effort that is optimal for the employer. In section 2, we are going to propose a theoretical framework where *ex post* evaluation interviews insure risk adverse agent against technological or market uncertainty.

In the French context, Crifo et al. (2004) observe that evaluation interviews are significantly more frequent when the employee is involved in collective work. The model of *ex post* evaluation interview is limited to give account of the function of evaluation interviews when employees work in teams, because individual performance appraisal become difficult when the output cannot be separated between the members of a team (Alchian and Demsetz, 1972). We are going to propose a model of *ex ante* individual evaluation interviews specific to the teamwork context, where evaluation comes first, before the constitution of teams and aims at fostering a team spirit. We conclude our theoretical approach by establishing some predictions about drivers and outcomes of individual evaluation interviews.

We assess empirically part of these predictions in section 3. Our empirical investigation rests on a matched employer / employee survey on organisational change and ICT use (COI). In the

labour force section of the survey, employees are asked whether they have been interviewed individually at least once in 1997. They also give information on work organisation, on personal characteristics and on outcomes. The business section of the survey gives a detailed set of firm level characteristics reflecting technological and organisational choices implemented in French manufacturing at the end of the 1990s. We use a propensity score methodology to evaluate the causal effect of individual evaluation interviews on effort, work overload and wage setting. In a last section, we conclude.

2. A theoretical analysis of evaluation interview

2.1. Definition of Evaluation Interview

Let us consider a Principal-Agent framework and let us define what an individual evaluation interview is. The theoretical determination of the optimal evaluation interview is a problem of its own that we are not going to solve in this paper. This would imply a rigorous assessment of the optimal¹ set of evaluation criteria. But we are going to leave this problem aside by simply assuming that the Principal determines (seeking advice from the Agent) a finite set S of criteria (characteristics) that he considers important for estimating the Agent's effort. Let us call $v^s \in \Xi^s$ the true level of agent's criterion s ($s = 1$ to S). We will assume that it is a private information namely that v^s is a random variable which realisation is only observed by the Agent (ie. The Principal does not observe it). According to the revelation principle we can restrict ourselves to the following direct mechanism $\Sigma^s = (\Xi^s; g)$ where g is a result function,

$$g : \Xi^s \rightarrow \Xi^s$$

$$v^s \mapsto g(v^s)$$

We call *individual evaluation interview*, the mechanism $\Sigma = (\Sigma^1, \dots, \Sigma^s, \dots, \Sigma^S; a)$ where the Σ^s ($s = 1, \dots, S$) are direct mechanisms and a is an aggregation function of marks resulting from the assessment of each criterion.

$$a : \prod_{s=1}^S \Xi^s \rightarrow M$$

$$(g(v^1), \dots, g(v^s)) \mapsto a(g(v^1), \dots, g(v^s))$$

Where M is a marks set.

1. In the sense that it minimizes the evaluation error.

This definition calls to two remarks. First in order to construct his own evaluation of the Agent's performance, the Principal may use the Agent's self-evaluations. Second, we assume the bounded rationality of the Principal, generating potential errors in his evaluation of the Agent's level of effort. Concerning this latter point, an interesting question from our point of view is to know whether it is theoretically possible to construct an incentive mechanism with individual evaluation interview when the Principal makes mistakes in his evaluation of the Agent's level of effort. Nonetheless in order to simplify we will assume that the probability of evaluation error is fixed and is common knowledge. To complete our comment, let us remark that since the purpose of the evaluation interview is to get a subjective evaluation of effort, then the Principal can directly take $M = \Theta$ (see the definition below).

Let us now explain our formal analysis concerning the role of individual evaluation interview². Since we want to take into account the production technology, we are going to divide our analysis into two parts. Firstly, we will analyse the function of individual evaluation interview in a productive context where work is "individualised", that is designed in such a way that tasks are not shared between employees. And secondly, we will analyse the function of individual evaluation interview in a productive context where work is organised in teams (i.e. when the implementation of a task is shared between employees, generating a need for cooperation and coordination).

2.2. The Role of Evaluation Interview in Individual Production

Concerning individual production, our theoretical framework is based on a Principal-Agent model with subjective evaluation of effort through an individual interview³. We consider a Principal-Agent framework à la Che and Yoo (2001) in which production requires only one task. This task is performed by the Agent who makes an effort decision unobservable by the Principal. Production, that is the outcome of the task, is a random variable X that can either succeed ($X=1$) or fail ($X=0$) giving respectively R or 0 payoffs to the Principal. The Agent's individual effort denoted K (K is a random variable from the Principal's stand point) belongs to the set $\Theta = \{0,1,2\}$ which is the set of *legally* possible levels of effort. In other words, the maximal level of effort that the Principal can legally incite the agent to supply is $K = 2$. However the general set of levels of effort is $\Theta_g = \{0,1,2,3,4,\dots,\overline{m}\}$. Let $\Pr(X=1|K=k) = q_k$ and $\Pr(X=0|K=k) = 1 - q_k$ respectively the conditional probability of success of the task given the Agent's level of effort k , and the conditional probability of failure of the task given the Agent's level of effort k . We will suppose that the Principal is risk-neutral, with a linear utility function

2. For more details, see Diaye et al. 2006.

3. Our model can therefore be considered as a sub-model of the one of MacLeod (2003) but not completely for four reasons. First in our analysis, the individual evaluation interview is based upon some precise common knowledge criteria (which are accepted by both parties -Principal and Agent). Second we assume that the Principal's subjective evaluation of the Agent's effort takes into account the Agent's self-reports on a set of criteria. Third the production technology matters. Finally, we do not address the question of the strategy-proofness or the Nash-implementability of the individual interview mechanism.

$b(r) = r$ and that the Agent is risk-averse with a utility function $U(r, k) = u(r) - v(k)$ where u is an increasing and concave function such that: $u(0) = 0$, $u(r) \geq 0 \quad \forall r \geq 0$; and the Agent's disutility function of effort $v(k) = e \times k$ where the unit of effort noted e is strictly positive. Moreover the Agent's reservation utility is equal to zero. Finally we assume⁴ that $1 > q_2 > q_1 > q_0 \geq 0$ and that $2q_1 \geq q_2$ (which guarantees that the participation constraint is fulfilled). The relationship between effort and production is an important feature of the production technology. As usual in the literature, we will consider two alternatives: production is either super-modular ($q_2 - q_1 \geq q_1 - q_0$) or sub-modular ($q_2 - q_1 \leq q_1 - q_0$). Broadly speaking when production is super-modular (respectively sub-modular), the return on effort is increasing (respectively decreasing) in the level of effort.

It is straightforward to see (and this is well known in the literature) that for each level of effort we have the following incentive wages:

$$\begin{array}{ll}
 & w_0^* = 0 \\
 \text{Sub-modularity:} & w_1^* = u^{-1}(e/(q_1 - q_0)) \quad w_2^* = u^{-1}(e/(q_2 - q_1)) \\
 \text{Super-modularity:} & w_1^* \text{ indeterminate} \quad w_2^* = u^{-1}(2e/(q_2 - q_0))
 \end{array}$$

When the production technology is super-modular the Principal will implement the mechanism $(w_2^*, k = 2)$, which is designed in such a way that the Agent plays the maximal level of effort. When the production technology is sub-modular and if income R is not high $(R \leq \frac{q_2 u^{-1}(e/(q_2 - q_1)) - q_1 u^{-1}(e/(q_1 - q_0))}{q_2 - q_1})$, then the Principal will ask for the level of effort $k = 1$. Otherwise, he will ask for a level of effort $k = 2$.

Hence in the classical mechanism, there is no need for the Principal to evaluate *ex post* the Agent's level of effort. So what is the role of individual evaluation interviews in the individual production? In order to understand our explanation, let us stress out two characteristics of the classical incentive mechanism (without individual evaluation interviews). The first one is that the Agent's wage will not directly depend on the level of effort, but instead on the success or the failure of the production. Of course, even if the Agent provides the maximal level of effort, the success of the production is not guaranteed. The second characteristic of classical incentive schemes is that if the Principal can choose between a sub-modular and a super-modular technology, he will always choose a super-modular one. The reasons are on the one hand because the Agent's wage $u^{-1}(2e/(q_2 - q_0))$ is lower than in the sub-modular technology case, and on the other hand because the probability of success of the task is then an increasing convex function of the level of effort. There is therefore a potential conflict between the preferences of the Principal and those of the Agent concerning the production technology. Basically, the Principal

4. Over $\Theta_g = \{0, 1, 2, 3, 4, \dots, \overline{m}\}$, we have $1 > q_{\overline{m}} > \dots > q_3 > q_2 > q_1 > q_0 \geq 0$.

always prefers a super-modular technology, while the Agent prefers a sub-modular one. Hence, although the incentive wage w_2^* respects the participation constraint, some agents may resist when the Principal implements a super-modular production technology. This resistance will be stronger when the firm moves from a sub-modular technology to a super-modular one with the same employees (indeed their wages will decrease⁵).

These two characteristics, from our point of view, may explain why it is difficult to implement a classical incentive scheme in real world. Indeed according to MacLeod and Parent (1999) and Prendergast (1999), in real world, very few firms use classical incentive schemes but instead a mechanism in which the employees' bonus depend on a subjective evaluation of their performance by the Principal.

On the contrary, in an incentive scheme with individual evaluation interview, the Agent will get this wage (even if the production is a failure) if he has been evaluated by the Principal as having provided an effort of level $k=2$. Moreover a mechanism with evaluation interview could mitigate the potential resistance of the agents concerning the implementation of super-modular technology, because the Agent then gets a higher wage than w_2^* , the classical incentive wage with a super-modular production technology.

Let I be the evaluation mark obtained by the Agent after the production process: $I \in \{0,1,2\} = \Theta$. If $I=0$ or 1 then the agent does not receive a premium. If $I=2$ then the agent receives a premium \bar{p} . Let us call p the premium variable. We thus have:

$$p = \begin{cases} \bar{p} & \text{if } I = 2 \\ 0 & \text{otherwise} \end{cases}$$

Of course, this mechanism must respect the participation and the incentive constraints. Moreover it must meet the following constraints:

- the principal expected benefit is at least equal to his expected benefit in the classical mechanism,
- the agent's expected utility is at least equal to his expected utility in the classical mechanism (because classical mechanism is the benchmark).

We assume here without loss of generality that the cost of an evaluation interview is equal to zero. Let us set $\gamma_k^i = \Pr(I=i | K=k)$, the probability that the Agent's level of effort were evaluated as being i when his true level of effort is k . Our bounded rationality assumption implies that the Principal make errors in his evaluation of the Agent's level of effort: $\gamma_k^i > 0 \quad \forall i, k$.

5. The decreasing of wage after an organizational change is not only theoretical. For instance in their empirical seminal paper, Hamilton et al. (2003) show that after a move from an individual production to a team one, high productivity workers take earnings losses when joining a team.

We will also assume that the probability of evaluating the Agent's effort equal to 2 when it is lower than 2 is independent from the true level of effort: $\gamma_{k'}^2 = \gamma^2, \forall k' \in \{0, 1\}$.

Further, we define an *evaluation system*, denoted E, as a profile

$$E = \left(\Theta = \{0, 1, 2\}, \{\gamma_k^i\}_{k,i} \right)$$

We will say that an *evaluation system* E is *efficient in detecting a level of effort* $k \in \Theta$ if the probability that the Agent's level of effort were evaluated as being k when his true level of effort is k is strictly higher than the same probability when his true level of effort is $k' < k$:

$$\gamma_k^k > \gamma_{k'}^k, \forall k' < k, k' \in \Theta.$$

We are going to discuss some claims highlighting different conditions under which the Principal implements individual evaluation interview in the case of individual production.

The *first claim* is that the efficiency of the Evaluation System for the level of effort 2 is a necessary condition to the implementation by the Principal of existence of individual evaluation interviews.

As a consequence, we are going to restrict ourselves to the class of Evaluation Systems which are efficient for the level of effort 2.

According to our *second claim*, when *the production technology is sub-modular*, then a necessary and sufficient condition for the implementation of an incentive mechanism with evaluation interview is $\frac{q_2}{q_1} \geq \frac{2\gamma_2^2}{\gamma_2^2 + \gamma_0^2}$: the marginal return from increasing the effort from level 1 to level 2 must compensate for the bounded rationality of the Principal.

The problem is that this necessary and sufficient condition will be difficult to fulfill⁶. Indeed when the technology is sub-modular ($q_1 - q_0 > q_2 - q_1$), the ratio $\frac{q_2}{q_1}$ is low. Therefore incentives mechanisms with evaluation interview will be rarely implemented when the production technology is sub-modular.

When *the production technology is super-modular*, then (*claim 3*) it is possible to construct an incentive mechanism with evaluation interview if and only if $\frac{q_2}{q_0} \geq \frac{\gamma_2^2}{\gamma_0^2}$: the relative increase in the probability of receiving the premium \bar{p} when the Agent switches from the level of effort 0 to the level of effort 2, has to be smaller than the relative increase in the probability of success of

6. Take for instance γ_2^2 close to 1 and γ_0^2 close to 0.

the task (which is for the principal the probability of receiving income R), when the Agent switches from the level of effort 0 to the level of effort 2.

This necessary and sufficient condition $\left(\frac{q_2}{q_0} \geq \frac{\gamma_2^2}{\gamma_0^2}\right)$ is not costly. For instance it is trivially satisfied when $q_0 = 0$. So incentive mechanisms with evaluation interviews have a high likelihood to be implemented when the production technology is super-modular.

Suppose now a super-modular production technology. Then (*claim 4*) the optimal contract in the incentive mechanism with evaluation interview is $\{\bar{p}, \gamma_2^2, \gamma'\}$ with:

- $\bar{p} = u^{-1}\left(\frac{2e}{\gamma_2^2 - \gamma_0^2}\right)$
- γ_2^2, γ' such as $u^{-1}\left[\frac{2e}{\gamma_2^2 - \gamma'}\right] \leq \frac{q_2}{\gamma_2^2 + 2\gamma'} u^{-1}\left[\frac{2e}{q_2 - q_0}\right]$

Two messages can be drawn from our fourth claim. According to the first message, the Agent's wage, $\bar{p} = u^{-1}\left(\frac{2e}{\gamma_2^2 - \gamma_0^2}\right)$, within the incentive mechanism with evaluation interview is higher than his wage in the classical incentive mechanism, $w_2^* = u^{-1}\left(\frac{2e}{q_2 - q_0}\right)$. Nonetheless the probability of receiving this wage is weaker: $\gamma_2^2 + \gamma_1^2 + \gamma_0^2 \leq q_2$.

In order to understand the second message, let us recall that in the classical optimal contract (with no evaluation interview) the Agent's wage is higher when the technology is sub-modular than when it is super-modular. If the firm moves from a sub-modular technology to a super-modular one with the same employees, the Principal may use evaluation interviews to mitigate potential resistance from the Agents: he will pay the same wage as in the case of a sub-modular technology with no evaluation interviews but he will reduce the probability for the Agent of getting this wage. Namely if the Principal wants to pay the Agent for the wage corresponding to the effort $k=2$ in the case of a sub-modular technology within the classical mechanism, then he will construct the incentive mechanism with evaluation interview in such a way that:

$$0 < \gamma' \leq \frac{1}{3} \left(\frac{q_2 u^{-1}\left(\frac{2e}{q_2 - q_0}\right)}{u^{-1}\left(\frac{2e}{q_2 - q_1}\right)^{sub}} - 2(q_2 - q_1)^{sub} \right)$$

$$\gamma_2^2 = 2(q_2 - q_1)^{sub} + \gamma'$$

Where $(q_2 - q_1)^{sub} = q_2 - q_1$ in the sub-modular technology case.

To summarise, in an incentive scheme with evaluation interview, the Agent gets his wage \bar{p} only if he has been evaluated by the Principal as playing an effort equal to 2; on the other hand this wage \bar{p} is greater than the one in the classical incentive scheme (however the probability of

getting \bar{p} is weaker). Therefore we expect that the incentive schemes with evaluation interview will attract people having a low disutility of effort.

Let us remind that $\Theta = \{0,1,2\}$ is the set of levels of effort legally possible but that the general set of levels of effort is $\Theta_g = \{0,1,2,3,4,\dots,\bar{m}\}$. Let us assume a continuum of agents but such that over $\Theta = \{0,1,2\}$ -the legal set of effort- agents have the same behavior with respect to their disutilities of effort. Therefore over $\Theta = \{0,1,2\}$ the Principal cannot offer a contract which depends on the Agent's type θ . Furthermore the Principal cannot construct a contract $(w_k, k \geq 3)$ because a court of justice may find from w_k the implicit level of effort (which is here illegal since it is higher than 2) that the Principal wants the Agent to provide. The disutility of effort is written as:

$$v_\theta(k) = \begin{cases} ke & \text{if } k \in \Theta = \{0,1,2\} \\ (2 + \theta k)e & \text{with } 0 \leq \theta \leq 1 \text{ if } k \in \{3,4,\dots,\bar{m}\} \end{cases}$$

Such a functional form means that given a level of effort $k \geq 3$, the smaller the type θ , the weaker the Agent's disutility of effort. Would an Agent of type θ increase his level of effort beyond the maximal legal level, without a monetary compensation⁷? In order to answer to this question, let us state that an evaluation system $E = (\Theta = \{0,1,2\}, \{\gamma_k^i\}_{k,i})$ includes an *informal detection system of a given illegal level of effort* $k' \geq 3$ if:

- 1) $\gamma_{k'}^i$ is well defined, and
 - 2) $\gamma_{k'}^i$ respects the following pseudo-monotony condition: $\gamma_{k'}^i \geq \gamma_{k'}^i$, $\forall k \in \{2,3,4,\dots,k'-1\}$.
- It is easy to see that when the evaluation system satisfies the above property then if the Agent increases his level of effort beyond the required level $k=2$, the probability of being detected (and thus of receiving the premium $\bar{p} = u^{-1}\left[\frac{2e}{\gamma_2^2 - \gamma_0^2}\right]$) as having provide a level of effort $k=2$, increases.

To illustrate, let us consider an agent of type $\theta=0$. Then if he provides an effort $k=\bar{m}$, he has the same effort disutility than when providing an effort $k=2$ with almost the certitude of getting the wage $\bar{p} = u^{-1}\left(\frac{2e}{\gamma_2^2 - \gamma_0^2}\right)$. Such an agent is rational when increasing his level of effort beyond the maximal legal level ($k=2$) because he increases his expected utility.

Hence an agent provides a level of effort k' superior strictly to the maximal legal level if the evaluation system $E = (\Theta = \{0,1,2\}, \{\gamma_k^i\}_{k,i})$ includes an informal detection system of the illegal level of effort k' .

7. Indeed, one can remark that the Agent's wage $\bar{p} = u^{-1}\left[\frac{2e}{\gamma_2^2 - \gamma_0^2}\right]$ does not change even if the Agent increases his effort beyond the required level.

According to our *claim 5*, assuming that the evaluation system $E = (\Theta = \{0,1,2\}, \{\gamma_k^i\}_{k,i})$ includes an informal detection system of the illegal level of effort k' , the Agent provides a level of effort at least equal to a given level k' (strictly greater than the maximal legal level) if and only if his type is $\theta < \delta(k')$ with $\delta(k') = \frac{2}{k} \frac{(\gamma_{k'}^2 - \gamma_k^2)}{(\gamma_2^2 - \gamma_0^2)}$ where $\bar{k} = k'1_{\{k=2\}} + (k'-k)1_{\{k \geq 3\}}$, $\forall k \in \{2, \dots, k'-1\}$.

The meaning of this claim is that if the evaluation system $E = (\Theta = \{0,1,2\}, \{\gamma_k^i\}_{k,i})$ includes an informal system of detection of the illegal effort then the incentive mechanism with evaluation interview always includes in addition to the normal incentive effect, a selection effect whose consequence is to attract the agents whose disutility of effort is the weaker. Of course we do not say that the incentives mechanism with evaluation interview only attracts individuals who will provide an effort greater than the maximal level required by the Principal. Indeed we have assumed that the optimal contract constructed by the Principal is based on the legal set of efforts $\Theta = \{0,1,2\}$. And over this set, the agents have the same behavior with respect to disutilities of effort. Nevertheless a simple reasoning shows that the selection effect will be higher in the case of production with evaluation interview than in the case without evaluation interview. To conclude, production structure with super-modular technology and evaluation interview will lead to an over-intensification of work in the sense that the Agents will provide efforts above the maximal effort required by the Principal. Of course since the probability q_k of success of the task is, over $\Theta_g = \{0,1,2,3,4, \dots, \bar{m}\}$, a monotone increasing function of the effort level k , then, assuming that the revenue R is high enough (i.e. $R \geq \bar{p} \left(\frac{\gamma_k^2 + \dots + \gamma_2^2}{q_k - q_2} \right)$), the selection effect also increases $q_k R - \bar{p}(\gamma_k^2 + \dots + \gamma_2^2 + \gamma_1^2 + \gamma_0^2)$: the *effective* expected profit of the Principal.

2.3. The Role of Evaluation Interview in Team Production

Concerning team production, even if the Principal faces the same problems as in individual production, the main issue (see Alchian and Demsetz, 1972 or Holmstrom, 1982) is free riding.

Therefore in team production, the role of evaluation interviews is not restricted to the *ex post* evaluation (through an individual interview) of the Agents' level of effort. Indeed an *ex post* evaluation of agents' level of effort will not prevent them from shirking (except if individual evaluation interviews as a mechanism can detect with probability equal to 1 any agent who shirks).

Thus we develop the following argument. The Principal prefers to implement a supermodular technology production. However in this case there are several Nash equilibria, among which the solution $(0,0)$ where the team produces no effort. Therefore in order to avoid the implementation of the equilibrium $(0,0)$, it must be the case that firms enforce coordination among the agents. The issue is not new and is well documented in the literature: coordination can be obtained using

non-monetary incentives or monetary incentives. A famous example of non-monetary incentives is the so-called "peer pressure" by Kandell and Lazear (1992). We show that *ex ante* (to the production) individual evaluation interviews belong to the class of non-monetary coordinating incentives and therefore play exactly the same role -concerning the implementation of the equilibrium (1,1) - as "peer pressure". However if implementing *ex ante* an evaluation interview solves the free-riding problem in team, the wage of the agents still depends on the conditional probability of success of the task. As in the individual production case, the Principal will implement an *ex post* evaluation interview which aims is to evaluate the level of effort and therefore to condition the wage not on the success of the task but on the evaluation of the Agents' effort.

We will consider the same framework as in the individual production case; however the task (which is the same as in the individual production case) is now performed by a team of two agents who make an effort decision unobservable by the principal. Our definition of team (borrowed from Alchian and Demsetz, 1972) is very restricted here since only a global signal X concerning production is available. In particular the Principal cannot calculate the contribution of each agent to the production. Finally, let us recall that in the individual production case, the set of levels of effort legally possible was $\Theta = \{0,1,2\}$. Therefore in the team production case (in order to be consistent with the case of individual production) it must be the case⁸ that from the point of view of the Principal, the individual effort of agent i ($=1,2$) denoted K_i belongs⁹ to the set $\{0,1\}$. The probability of success of the task given the level of effort of the agents is :

$$\Pr(X = 1 | K_i = k_1, K_j = k_2) = p_{k_1 k_2} \quad ; \quad i = 1, 2; j = 1, 2, i \neq j .$$

We assume also that: $p_{11} > p_{01} = p_{10} > p_{00} \geq 0$.

It is usual in the literature to state the following two properties of the team production technology. The *first one* is of course the super-modularity ($p_{11} - p_{10} \geq p_{10} - p_{00}$) or sub-modularity ($p_{11} - p_{10} \leq p_{10} - p_{00}$). The *other property* is horizontal technological interdependence ($\Delta_k = p_{k1} - p_{k0}$) which measures the increase in the probability of success (i.e. in the productivity) when one agent increases his effort level, for a given effort level of the other team member. When they are positive, such interdependencies capture the fact that teamwork makes an employee's productivity more dependent on the effort expanded by his co-workers.

In team production, only a global signal is available for the Principal. Hence, he can only propose contracts with symmetric effort levels. $w_{k_1 k_2}^*$ is the optimal wage for effort levels k_1 et k_2 required to agents 1 and 2 with $w_{00}^* = 0$ and $w_{11}^* = u^{-1}\left(\frac{e}{p_{11} - p_{01}}\right)$. In team production, after signing the contract proposed by the principal, agents make their effort decision in the context of a coordination game. It is easy to see that this game leads to a unique Nash equilibrium (1,1) when $\Delta_1 < \Delta_0$ and to two Nash equilibria (0,0) and (1,1) when $\Delta_1 \geq \Delta_0$.

8. See Che and Yoo, 2001.

9. Indeed the required total amount of effort for one task in team production is therefore the same as in the individual production case.

However w_{11}^* is a decreasing function of Δ_1 . Hence ceteris paribus, firms will always prefer organisational forms characterized by both high horizontal interdependencies and super-modularity ($\Delta_1 \geq \Delta_0$)¹⁰. The problem is that in the case of super-modularity, the equilibrium (1,1) is no longer unique. Therefore in order to avoid the implementation of the equilibrium (0,0), it must be the case that firms enforce coordination¹¹ among the agents. We will now show that *ex ante* individual interviews belong to the class of non-monetary coordinating incentives (and therefore play exactly the same role -concerning the implementation of the equilibrium (1,1) - as "peer pressure"). Indeed, faced with such a moral hazard problem within his team, each agent ignores whether his co-worker shirks or not. Each individual hence forms beliefs over his partner's strategy, and chooses an effort level given these beliefs. As argued by Osborne and Rubinstein (1994, Chapter 3, section 3.2) the probabilities ($\Pr(k_i = 1)$ and $\Pr(k_i = 0)$) may be interpreted as the players' beliefs about their peer's behavior. Unfortunately, it is easy to see that the only mixed strategies Nash equilibria are degenerated: ((1,0),(1,0)) and ((0,1),(0,1)). The meaning is that both agents make an effort decision $k_i = 1$ ($i = 1, 2$) only if they are sure that their partner adopts the same strategy. If the principal wants agents to coordinate on the Pareto-optimal equilibrium, he must design organisational devices affecting the probabilities associated with each strategy. This is the reason why we argue in claim 6 that evaluation interviews represent a mean to generate an *ex ante* signal towards the other agents to reinforce their beliefs about the team spirit¹². An important difficulty lies in the fact that the outcome of an evaluation interview is only known by the principal and the interviewed agent. It is therefore not public. In turn, it seems hard to figure out how the evaluation interview of an agent i would generate a signal (regarding himself) towards the other agents j , $j \neq i$. Though this may seem counter-intuitive at first sight, we will show that the individual evaluation interview is such that it does in fact make public the outcome of the interview. We assume that the interview is run before production takes place and that agent i 's type $t_i \in \Lambda_i = \Lambda = \{H, L\}$ is a private information where H = "has team spirit" and L = "has no team spirit". The goal of the interview is of course to make agents truthfully reveal their types.

Let now define the following¹³ mechanism *CDG* :

10. Indeed let Δ_1^a such that $\Delta_1^a < \Delta_0$ and Δ_1^b such that $\Delta_1^b \geq \Delta_0$. We have therefore: $w_{11}^*(\Delta_1^a) > w_{11}^*(\Delta_1^b)$.

11. By coordination within the team, we mean any instrument which leads to the implementation of the Nash Pareto-Optimal equilibrium (1,1).

12. In the sense of Alchian and Demsetz (1972), an agent has team spirit (or loyal) if he does not shirk when working in a team.

13. If the principal does not know the proportion of agents who have team spirit, then this proportion is a random variable for him.

Let us call it N and let us suppose that the principal knows its support $Sp = \left[\frac{n_{\inf}}{n}, \frac{n_{\sup}}{n} \right]$, with of course

$n_{\inf} \geq 2$. If the principal takes *InfSp* as the proportion of agents who have team spirit (that is if the principal is prudent) then the mechanism *CDG* is still strategy-proof.

- 1) The interview is run over a population of n individuals, $n \geq 2$.
- 2) The Principal knows (whithout loss of generality) the proportion $\frac{n_0}{n}$ of agents who have team spirit, where $n_0 \geq 2$.
- 3) The Principal runs the interview and if he observes that the number of individuals who declare to be of type H (labeled \hat{n}_0) is strictly higher than n_0 , then he knows that at least one individual (among the \hat{n}_0) lies.
In such a case, the Principal stops the process : *No team*.
- 4) If $\hat{n}_0 \leq n_0$, the Principal randomly selects two agents among the \hat{n}_0 who declare to be of type H .
- 5) Each interviewed individual only receives the outcome of his own interview.

It is easy to see that (*claim 6*) if the mechanism¹⁴ CDG is common knowledge then *it is strategy-proof*.

This claim implies that when the Principal forms the team, each agent within the team infers that the other team member has a team spirit (CDG being common knowledge), even though he does not observe the outcome of the other agent's interview. Moreover, it follows directly from the definition of team spirit that the agents do believe that the signal t_i is perfectly correlated with the effort variable of agent i , k_i , so that when agent i observes $\theta_j = H$ (respectively $\theta_j = L$), $i \neq j$, he believes that $k_j = 1$ (respectively $k_j = 0$) and it is then optimal to decide $k_i = 1$ (respectively $k_i = 0$). Thus when the mechanism CDG is implemented, the unique Nash correlated equilibrium is (1,1). Our claim 6 is somewhat close to a result by Prat (2002) showing that when there is a positive complementarity between workers in a team then this team should be composed of agents of the same type.

If implementing *ex ante* (*ex ante* to the production) an evaluation interview solves the free-riding problem in team by selecting agents who have team-spirit, the wage of these agents still depend to the probability of success of the task p_{11} . As in the individual production case, the Principal can implement an *ex post* evaluation interview which aims is to evaluate the level of effort and therefore to condition the wage not on the success of the task but on the evaluation of the Agents' effort.

Let γ_0^1 (respectively γ_1^1) be the probability that the Agent's level of effort were evaluated as

14. Of course the mechanism CDG can select people having the same demographic characteristics (gender, age, race,...) either because team spirit is objectively correlated with the agents' individual characteristics or because the principal and/or the agents have some prior beliefs concerning the demographic characteristics of people having team spirit.

being 1 while his true level of effort is 0 (respectively 1). Let us assume that γ_0^1 and γ_1^1 are strictly positive; and that the evaluation system is efficient in detecting a level of effort 1 (that is to say $\gamma_0^1 < \gamma_1^1$). Then the Principal will maximise his profit $p_{11}R - 2(\gamma_1^1 + \gamma_0^1)\bar{p}$, where \bar{p} is the Agent's wage when he has been evaluated as having provided the level of effort 1; under the participation and incentives constraints and also under the constraint that his (the Principal) expected benefit is at least equal to his expected benefit in the mechanism with only an *ex ante* individual evaluation and under the constraint that the agents' expected utility is at least equal to their expected utility in the mechanism with only an *ex ante* individual evaluation.

The resolution of this program implies that the Principal can implement two kind of policy: either \bar{p} (the agent's wage in the incentive mechanism with *ex ante* and *ex post* evaluations) $\geq w_{11}^*$ (the agent's wage in the incentive mechanism with only an *ex ante* evaluation) and $\gamma_1^1 + \gamma_0^1 \leq p_{11}$ (the probability of getting \bar{p}) $\leq p_{11}$ (the probability of getting w_{11}^*) or $\bar{p} < w_{11}^*$ and $\gamma_1^1 + \gamma_0^1 > p_{11}$.

However we think that the Principal will implement a mechanism in which $\bar{p} \geq w_{11}^*$ and $\gamma_1^1 + \gamma_0^1 \leq p_{11}$ instead of a mechanism with $\bar{p} < w_{11}^*$ and $\gamma_1^1 + \gamma_0^1 > p_{11}$. The reason is that in the former, the effect of the selection effect on the Principal's effective expected profit will be higher. Indeed the smaller is $\gamma_1^1 + \gamma_0^1$ (the probability of getting the wage \bar{p}) the higher will be the agent's effort in order to increase the probability $\sum_{i=0}^k \gamma_i^1$ to be evaluated playing an effort equal to 1. We can also remark, concerning the selection effect, that if $\gamma_2^1 - \gamma_1^1 \geq \gamma_1^1 - \gamma_0^1$ (that is if the evaluation technology is super-modular) then all the agents will play the level of effort $k = 2$ (instead of the required $k = 1$)¹⁵. More generally, if the evaluation system includes an informal system of detection system of an illegal level of work ($k \geq 3$) then the agents of type $\theta < \frac{(\gamma_k^1 - \gamma_1^1)\mu(\bar{p}) - e}{ke}$ will play an effort $k \geq 3$.

To conclude, in team production evaluation interviews play two roles. On the one hand, they play the same role as in individual production, and on the other hand evaluation interviews is used by the firms as a non-monetary coordinating device which permits to solve the free riding problem within team, by selecting some individuals having the same characteristics correlated to team spirit. Among these two roles we think that the most important is actually the last one.

2.4. Some testable predictions

We can derive some predictions from our theoretical analysis.

15. And if the revenue R is high enough (*i.e.* $R \geq \frac{\gamma_2^1 \bar{p}}{p_{22} - p_{11}}$) then this overintensification of work will increase the Principal's effective expected profit.

- 1) In individual or team production, evaluation interviews are associated with super-modular production technology. More precisely, while the conditions to implement evaluation interviews are difficult to fulfil when the production technology is sub-modular, evaluation interviews are easy to implement when the production technology is super-modular. The diffusion of evaluation practices are thus a symptom of the diffusion of super-modular technologies. Collective work practices could also reflect the diffusion of super-modular technology because they allow to reach higher level of effort by sharing the burden collectively, mitigating the pressure weighing on individual workers.
- 2) In individual production, evaluation interviews are mainly devoted to the *ex post* evaluation of the employees effort. In team production, evaluation interviews are of course also devoted to the *ex post* evaluation of the employees effort; nevertheless they play another important *ex ante* role which is to solve the free riding problem within teams. As a consequence, we expect that in team production, employees will be more evaluated (than employees working in individual production) and that their evaluation devices will be more complex.
- 3) From the employees' point of view, evaluation interviews are a kind of insurance. Indeed in an incentive scheme with evaluation interviews, on the contrary to the classical incentive scheme, the Agent will get this wage even if the production is a failure, if he has been evaluated by the firm as having provided the required level of effort of level. Moreover in the case of team production, evaluation interviews solve the free riding problem within teams.
- 4) The agents' level of effort in incentive scheme with evaluation interviews is higher than the level of effort in the classical incentive scheme.
- 5) In individual production, a mechanism with evaluation interview always includes in addition to the standard incentive effect, a high selection effect whose consequence is to attract the agents whose disutility of effort is the weaker.
- 6) In team production, a mechanism with evaluation interview includes in addition to the standard incentive effect, two selection effects linked respectively to team spirit and disutility of effort.
- 7) The impact on the level of effort of the selection effect concerning disutility of effort is higher in individual production than in team production.
- 8) Evaluation Interviews in a context of super-modular technology lead to an overintensification of work in the sense that the employees' level of effort will be higher than the one "required" by the firm. The reason is the selection effect regarding disutility of effort.
- 9) In firms which implement a scheme with evaluation interviews, the employees wage is higher than the one in a classical incentive scheme.
- 10) In a scheme with evaluation interviews, the wage depends on the employee's evaluation mark which is itself based upon some precise common knowledge criteria (which are accepted by both parties -Firm and Employees). Hence evaluation interviews will be associated with a better objectivation of wage from the point of view of employees.

Of course our explanations of subjective evaluation through evaluation interview are not exclusive. Indeed other explanations exist in sociology, theories of organisations or industrial relations. For instance, evaluation interviews may contribute to elaborate the formalization of work organization, or they could deter social unrest within organizations in which the dispute potential is important¹⁶.

3. Empirical tests

We are going to use a matched employer / employee survey, the survey on "computerisation and organisational change" (COI) to assess our theory of evaluation interviews. The information we have in the survey will not allow us to test all the predictions from our model. However, from the employee section of the survey, we have some information on the characteristics of work (whether individual or collective), on evaluation interviews, on effort and on wages. This will allow us to cover the main features underlined by our theoretical model. More precisely, we will be able to test whether evaluation interviews lead to higher levels of effort than classical incentive schemes (prediction 4). Our estimation strategy will also allow us to assess the existence of a selection effect associated with the implementation of evaluation interviews in individual and collective work organisations (predictions 5, 6 and 7). Furthermore, using measures of work overload, we will check whether evaluation interviews drive workers towards an excessive work intensity leading to inefficiencies. This is a possible consequence of our prediction 8 on overintensification. It is important to test this implication because work overload is a major factor of stress and has long term implication on the health of the workforce, especially in a context of aging. Finally we will be able to test our predictions on wage differentials and on the employees' knowledge of the rules driving wage setting between the scheme with evaluation interviews and the classical incentive scheme (predictions 9 and 10).

3.1. The Data

The COI survey has been conducted at the end of 1997 by the French public statistical system¹⁷. We are going to work on a representative sample of manufacturing firms with more than 50 employees and on a sample of randomly selected employees within these firms. In matched employer/employee surveys, the budget constraint implies a trade-off between trying to capture the diversity of firms and trying to capture the diversity of the work force within firms. By choosing to interview small sample of employees (1, 2 or 3) within each firm, COI chooses to favour the diversity of firms. As interviewed employees have at least one year of seniority within the firm, they belong to its core workforce.

16. Faced with the possibility of expressing themselves during interviews, employees would be less incited to contest management.

17. The conception and coordination of the COI survey has been directed by the Centre d'Etudes de l'Emploi. The survey has been carried out in a consortium involving the Ministry of labour (DARES), the Ministry of Industry (SESSI), the Ministry of Agriculture (SCEES) and the National Institute of Statistics and Economic Studies (INSEE). It benefited from very high response rates: 82% for employers and 75% for the employees. For a detailed description of the survey, see Greenan and Hamon-Cholet (2001).

In the full sample of the labour force section of the survey, there are 4295 employees. However in our analysis we do not take into account employees with supervision activities (1214 individuals) or employees working part time (177 individuals). Indeed the former combine a position of Principal and of Agent that we have not investigated theoretically, while part time leads to badly measured effort and wages. We obtain a sub-sample of 2904 employees.

The available information on the practice of individual evaluation interviews stems from the following question: Do you have at least one evaluation interview per year? (yes / no). Because of their seniority in the firm, we know that all interviewed employees had the opportunity of being evaluated at least once.

The labour force section of the COI survey describes in detail work organisation. It includes a whole set of questions capturing whether work is structured around group activities. From these questions we build up five different measures of interaction between employees in the course of the work process: being part of a team, time spent in teamwork, intensity of communication with other workers, level of support from other workers, participation into meetings (see appendix 1 for detailed questions). These five measure are positively correlated, with correlations ranging between 0.04 (intensity of communication with time spent in team work) and 0.18 (being part of a team and level of support from other workers). Thus they measure different dimension of collective work. We derive from these five measures a synthetic binary indicator of collective work. When it takes the value one, the employee is considered as being a "collective" worker, when it takes the value 0, he is considered as being an individual worker. According to this variable our sample of employees breaks down into 1537 individual workers and 1367 collective workers.

Table 1 gives the distribution of individual evaluation interviews according to our synthetic binary indicator of collective work. 37.2% of the employees have been interviewed at least once in 1997. Evaluation interviews are positively correlated with collective work: 47% of collective workers have been evaluated against 29% of individual employees.

Table 1: Evaluation among individual workers and collective workers

	Individual Workers	Collective Workers
Evaluation		
Yes	445 (29 %)	637 (47 %)
No	1092 (71 %) ^a	730 (53 %)
Total	1537(53 %)	1367(47 %)

(a) : Percentage with respect to the analysed subsample of 2904 employees.

The COI survey also measures different effort indicators. Productive effort is measured through two questions indicating if the employee works longer than the usual hours some days or some weeks. Productive effort is considered as very high if the employee sometimes increases hours worked for personal reasons, as high if he or she sometimes increase hours worked in response to the firm's demand and as low if longer hours never happen. According to these three situations, the productive effort indicator respectively takes a value of 2, 1 or 0. The cognitive effort indicator is a binary variable indicating if the employee makes propositions to improve his or her work stations, the production process or the machines. It measures an involvement into collective knowledge building about the productive activity, allowing continuous improvement of the

production process.

Two additional measures are included in the analysis to identify if effort is going beyond reasonable levels, creating an overload that could be detrimental for work efficiency and for the employee's health. A first variable indicates how often an employee has to hurry in the course of his or her work. Four states are taken into account: hurrying almost all the time, hurrying for one quarter of the time or more, hurrying for less than a quarter of the time and never. The hurry variable respectively takes the value 4, 3, 2 and 1 according to the intensity of the pressure. Work overload is also measured through a binary indicator telling whether the employee often has to interrupt one task to carry out another urgent and non anticipated one.

Finally, we measure the employee's annualised net wage in euros. As it comes from an administrative data file used to compute social contributions, it is precisely measured and includes all bonuses, taxed allowances and compensations in kind. We also build up an indicator of the employees' ability to predict their wage. It rests on a question about the elements that have a big influence on the employee's wage or promotion, followed by a list of eight items. We compute the ratio of the number of yes to the list of items, on the number of yes and no, which gives an indicator taking its value between 0 and 1. 0 means that the employee has no idea of how to increase his or her wage and/or chance of promotion, 1 means that the employee knows that he or she can improve his or her situation and is aware of what to do to obtain this outcome.

3.2. Estimation strategy

We want to measure the impact of evaluation interviews on effort, work overload and wages, but we know, from our theoretical model that evaluation interviews induce a selection process. Employees with a low disutility of effort and, in the case on team work, with a team spirit are going to be attracted by jobs where evaluation interviews are conducted periodically. A possible way to measure outcomes related to evaluation interviews, taking into account the selection effect is to consider evaluation interviews as treatments and to apply a propensity score method to match each treated individual with a non-treated individual with the same characteristics in order to turn our non experimental data into a quasi experiment.

A simple way to test the predictions of the model is to consider evaluation interviews as treatments and to evaluate the effect of this treatment on the chosen variables for measuring effort, wages, and beliefs about wages. More precisely, let t a dummy variable equal to 1 if the employee declares being evaluated and 0 otherwise, then 3 quantities are of interest to us. The first is the average treatment effect over the whole population, written C ; the second is the average treatment effect over the treated individuals, written C_1 ; and the third is the average treatment effect over the non-treated individuals, written C_0 . More precisely, let Y the chosen variables for measuring effort, wages, and beliefs about wages. Then C measures the variation of Y that would be observed if the whole population was treated; C_1 is an evaluation of the effect of the treatment in the usual sense since it concerns the treated population; and C_0 is a prospective evaluation in the sense that it measures what would happen if the non-treated population was treated. We have:

$$C = E[Y_1 - Y_0]$$

$$C_1 = E[Y_1 - Y_0 | t = 1]$$

$$C_0 = E[Y_1 - Y_0 | t = 0]$$

Where Y_1 is the observed value of Y that results when receiving treatment (that is when being evaluated), Y_0 is the observed value of Y that results when do not receiving treatment (that is when do not being evaluated), and $E[.]$ denotes expectation in the population. Intuitively, an estimate of an average treatment effect could be the difference between the average of Y over the population of treated individuals and its average over the population of non-treated individuals, that is,

$$\bar{Y}_1 - \bar{Y}_0$$

Where \bar{Y}_1 and \bar{Y}_0 are respectively the average of Y for treated (evaluated employees) and the non-treated (non evaluated employees).

However, broadly speaking the main problem when evaluating the effect of a treatment is that for each individual we only observe

$$Y = t \times Y_1 + (1-t) \times Y_0 = \begin{cases} Y_1 & \text{if } t = 1 \\ Y_0 & \text{if } t = 0 \end{cases}$$

Then it can be shown that the average difference between treated and non-treated individuals can be the cause of a selection bias since the data does not result from a randomised experiment. And when testing evaluation effects (on effort, overload and wages) there is a need to control for naturally occurring systematic differences in background characteristics between the treated population and the non-treated population, which would not occur in the context of a randomised experiment. Moreover, according to predictions 5 and 6, individual evaluation interview affects employees' effort through a selection effect associated to disutility and/or to team spirit an incentive effect which in our case is estimated by the average treatment (evaluation) effect. Therefore in order to estimate the average treatment (evaluation) effect it is also necessary to control for the selection bias due to disutility. Although, it seems difficult to control "directly" for this selection effect because disutility and/or team spirit are not observable characteristics, we can assume that they are grounded on observable background characteristics of the employee and of the employer, and hence controlling for them allows to control for the selection.

We will discuss in the next section the background characteristics we will take into account to estimate the effect of individual evaluation interview. We choose we choose to use the propensity score methodology introduced by Rosenbaum and Rubin (1983). This method reduces the entire collection of background characteristics to a single composite characteristic that appropriately summarises the collection. Propensity score technology allows to correct the selection bias by matching individuals according to their propensity score which is the estimated probability of receiving the treatment (of being evaluated) given background characteristics. We are going to use a non parametric *Kernel matching estimator* proposed by Heckman, Ichimura, and Todd (1997, 1998) which under some regularity assumptions is convergent and asymptotically normal.

3.3. Determinants of individual evaluation interviews

The first step of the propensity score method is to analyse the determinants of evaluation interviews taking into account background characteristics that influence the employee's probability of receiving a periodical evaluation interview and the three categories of outcomes we consider: effort, work overload and wages.

In this step, it is very important to take into account individual effects as well as contextual effects. As we have pointed out in our theoretical model personal characteristics of the employee like team spirit or disutility of effort are going to play a crucial role in influencing both the chances of being evaluated and the outcomes we consider. These characteristics are not directly observable, but we are going to take into account observables that are possibly correlated with them: gender, age, seniority, education level and occupation. It is clear that these personal characteristics have impacts on effort levels, work overload and wages.

The fact that our employee sample is matched with a survey describing the characteristics of firms is an important advantage in our estimation strategy. Our theoretical model has stressed that the production technology plays a role in the diffusion of evaluation interviews. A super-modular technology is more favourable than a sub-modular technology. In order to control for the technology, we are going to include in the regression size and sector dummies. Stemming from an employer database, information on size and sector is much more precise than the information usually included in labour force surveys. We also include a measure of the firm's computerisation intensity. We choose to build up a variable describing the intensity of numerical data transfers within and outside the firm. Moreover, evaluation interviews could be complementary to other organisational practices and these practices could also have an influence on outcomes. Eight new organisational practices are considered in the logistic regression: quality certification, total quality management, methods to analyse products and processes (value analysis, functional analysis, AMDEC), total productive maintenance (TPM), organisation in profit center, formal in house customer/supplier contracts, system of just in time delivery, system of just in time production. We also detail different teamwork practices: self managed teams, problem solving groups and project teams. Finally, we take into account the evolution of the number of hierarchical layers in the firm and variables indicating difficulties connected with the implementation of organisational changes.

Appendix 3 presents the parameters estimated of the logistic models explaining individual evaluation interviews for individual workers and for collective workers. In the case of individual workers we find that employee characteristics have higher explanatory power than employer characteristics. More precisely, male workers in executive or middle management positions with either low seniority (1 or two years) or intermediate seniority (7 to 10 years) have a higher probability of being evaluated. We have to keep in mind that even though some of the interviewed workers have management positions, they have no formal hierarchical authority as they declare no subordinates. Among the employer characteristics, the only variables with significant influence are size, with of positive impact of the highest size cluster, sector, with a positive impact of five sectors (pharmaceutical, perfumes, and cleaning products; chemicals, rubber, and plastic products; electrical and electronic equipment; electrical and electronic components; and shipbuilding, aircraft, and railway) and quality certification (ISO 9001, ISO 9002 and EAQF).

By contrast, in the case of collective workers, employer characteristics tend to explain more than employee characteristics. Indeed for team workers the only personal characteristic that influences the probability of being evaluated is the level of education: a second or a third level of education is associated with a coefficient that is positive and significant. On the employer side, size, sector, computer intensity, use of new organisational devices and use of teamwork have a significant impact on the probability of being evaluated. Employers with medium size (between 100 and 999 employees) and belonging to pharmaceutical, perfumes, and cleaning products or to chemicals, rubber, and plastic products use evaluation interviews more frequently. Employers from printing, press and publishing and shipbuilding, aircraft and railways have a lower probability of being interviewed. The intensity of computerisation favours evaluation interviews of collective workers, as well as quality certification and total productive maintenance. On the opposite, employers using just in time delivery are less oriented towards evaluation interviews for collective workers. Having a non marginal fraction of production workers in problem solving groups favours evaluation interviews, while having a small fraction of non production workers participating in self managed teams and having management involved in project teams has a negative impact on evaluation interviews. In total, evaluation interviews for collective workers seem complementary with information technologies and new organisational practices. These managerial tools could support a super modular production technology, where the employer has a preference for higher levels of effort.

3.4. Observing the outcomes of individual evaluation

We are now going to discuss the matching evaluation of the effect of individual evaluation interviews on individual and collective workers on effort (table 2), work overload (table 3) and wages (table 4). In each table, we first compute as a benchmark the average outcome for individual and collective workers. Second we compute the average difference in outcome between workers that have been individually evaluated and workers that have not been evaluated. This estimator is often designated as the naï ve estimator of the treatment effect. Then we compute the three causal effects: the effect on the treated (C1), the effect on the non-treated (C0) and the global effect (C). The first effect is the matching evaluation strictly speaking, the second one represents the effect that evaluation interviews would have if they were implemented on the non evaluated population of workers and the last one is the effect that would be obtained if evaluation interviews were extended to the entire population.

a. Effort

We observe higher levels of productive and cognitive efforts when work is collective rather than individual (table 2). This was not entirely expected because our model underlined that one of the advantage of collective work was to share the burden of higher levels of effort between workers. However, other effects might play a role here. Our theoretical model, as well as, the analysis of determinants of evaluation interviews suggests that collective work is positively correlated with super modular production technologies. Another explanation could lie in synergy and peer pressure effects connected with collective work.

Table 2: Individual evaluation interviews and Effort

	INDIVIDUAL WORKERS ^a	COLLECTIVE WORKERS ^b
• PRODUCTIVE EFFORT ^d		
Average productive effort	0.564	0.720
Average difference E/NE	0.127 *** ^c	0.092 ***
Effect on the treated (C1)	0.084 **	0.120 **
Effect on the non-treated (Co)	0.093 **	0.100 **
Global effect (C)	0.091 **	0.110 **
• COGNITIVE EFFORT ^d		
Average cognitive effort	0.507	0.722
Average difference E/NE	0.143 ***	0.140 ***
Effect on the treated (C1)	0.099 **	0.110 **
Effect on the non-treated (Co)	0.120 **	0.110 **
Global effect (C)	0.114 **	0.110 **

(a) : The standard deviation of the treatment effect is computed using bootstrap with 300 simulations. The characteristics of the support over 300 simulations are : min=1352; max=1501; mean=1426.48

(b) : The standard deviation of the treatment effect is computed using bootstrap with 300 simulations. The characteristics of the support over 300 simulations are: min=1124; max=1304; mean=1229.03

(c) *** p-value<0.01 , ** 0.01≥ p-value<0.05 , * 0.05 ≥ p-value < 0.1 .

(d) : See B. of appendix 1 for a description of these variables.

As predicted by our model, we observe that the level of effort whether productive or cognitive is higher when workers are individually evaluated than in the classical incentive scheme (prediction 4).

The causal treatment effect on productive effort is stronger for individual workers than for collective workers. And the selection effect is of opposite sign. Individual workers displaying higher level of effort are selected in the population of evaluated workers, when they are selected out in the case of collective work. This result corroborates our prediction 7 although our model gives no specific clue to understand our surprising result on collective workers. The extension of evaluation interviews to the whole population of collective workers would consequently increase productive effort, although it is already high in this case.

The observed effects on cognitive effort are more straightforward. Evaluation interviews similarly affect cognitive effort for individual and collective workers: they increase by 14% the propensity to make propositions for improving the production process. In the case of cognitive effort, the selection effect has identic sign among individual and collective workers, but it is stronger in the first case, as stated in prediction 7.

b. Work overload

Individual and collective workers work with a similar time pressure: the average of hurry is 2.67 in the first case, 2.64 in the second (table 3), indicating that workers have to hurry a little more

than a quarter of their time. Our second indicator of work overload is higher for collective workers: 65% of collective workers experience task interruptions in the course of their work whereas 53% of individual workers face interruptions.

Table 3: Individual evaluation interviews and Work overload

	INDIVIDUAL WORKERS ^a	COLLECTIVE WORKERS ^b
● HURRY ^d		
Average overload	2.666	2.640
Average difference E/NE	-0.143 ** ^c	-0.110 *
Effect on the treated (C1)	-0.142 (ns)	-0.108 (ns)
Effect on the non-treated (Co)	-0.189 **	-0.073 (ns)
Global effect (C)	-0.176 **	-0.089 (ns)
● INTERRUPT ^d		
Average overload	0.526	0.650
Average difference E/NE	-0.053 *	0.009 (ns)
Effect on the treated (C1)	-0.065 **	0.002 (ns)
Effect on the non-treated (Co)	-0.066 **	-0.003 (ns)
Global effect (C)	-0.066 **	-0.000(ns)

(a) : The standard deviation of the treatment effect is computed using bootstrap with 300 simulations. The characteristics of the support over 300 simulations are : min=1352; max=1501; mean=1426.48

(b) : The standard deviation of the treatment effect is computed using bootstrap with 300 simulations. The characteristics of the support over 300 simulations are: min=1124; max=1304; mean=1229.03

(c) : *** p-value<0.01 , ** 0.01≥ p-value<0.05 , * 0.05 ≥ p-value < 0.1 .

(d) : See C. of appendix 1 for a description of these variables.

However, it is in the case of individual workers that evaluation interviews have a significant impact as it appears to mitigate work overload. Individual workers that are periodically evaluated work under lower time pressure and are less exposed to task interruptions. In the case of time pressure, the selection effect seems to play an important role as the causal effect on the treated is not significant. But evaluation interviews also seem to have a protective effect on their own because the effect on the non-treated is negative, significant and stronger than the naïve estimator. Individual workers who have been selected out from evaluation interviews would benefit from their implementation. As far as task interruptions are concerned, the protective effect of evaluation interviews is not explained by a selection effect, it is a pure outcome of this managerial device.

Evaluation interviews do not protect collective workers from work overload, but they do not increase their risk of exposition either. It is also an interesting result, knowing that collective workers produce higher levels of productive and cognitive efforts.

These results could be indirect evidence of prediction 8. Evaluation interviews in a context of super-modular technology lead to an overintensification of work, but not to work overload. On the contrary, they seem to mitigate work overload, either through a selection effect as described

in our theoretical model, or through a pure effect that we have not investigated theoretically.

c. Wage setting

Collective workers earn more, on average than individual workers (table 4). We also observe that on average evaluated employees earn more than employees in a classical incentive scheme, confirming prediction 9. These monetary gains are higher for collective than for individual workers: 1925 euros per year on average against 1654 euros per year. For individual workers, this difference is entirely explained by the selection effect: the causal effects on the treated is not significantly different from zero and the causal effect on the non-treated is also non significant. Contrary to individual workers, the monetary gain of collective workers is only slightly lower when selection is taken into account: the gain falls from 1925 euros to 1310 euros if we consider the causal effect on the treated, to 1062 if we consider the causal effect on the non-treated, and to 1174 if we consider the global effect.

Table 4: Individual evaluation interviews and wage setting

	INDIVIDUAL WORKERS ^a	COLLECTIVE WORKERS ^b
• ANNUALISED NET WAGE (IN EUROS) ^d		
Average net wage	15003	16586
Average difference E/NE	1654 *** ^c	1925 ***
Effect on the treated (C1)	198 (ns)	1310 **
Effect on the non-treated (Co)	275 (ns)	1062 **
Global effect (C)	253 (ns)	1174 **
• EMPLOYEE'S ABILITY TO PREDICT HIS OR HER WAGE ^d		
Average ability to predict	0.491	0.597
Average difference E/NE	0.164 ***	0.136 ***
Effect on the treated (C1)	0.145 ***	0.110 ***
Effect on the non-treated (Co)	0.147 ***	0.100 ***
Global effect (C)	0.146 ***	0.100 ***

(a) : The standard deviation of the treatment effect is computed using bootstrap with 300 simulations. The characteristics of the support over 300 simulations are : min=1352; max=1501; mean=1426.48

(b) : The standard deviation of the treatment effect is computed using bootstrap with 300 simulations. The characteristics of the support over 300 simulations are: min=1124; max=1304; mean=1229.03

(c) : *** p-value<0.01 , ** 0.01≥ p-value<0.05 , * 0.05 ≥ p-value < 0.1 .

(d) : See D. and E. of appendix 1 for a description of these variables.

Concerning the *employee's ability to predict his or her wage* we first note that this ability is greater on average for collective workers than for individual workers and in both cases the

average difference between evaluated and non evaluated workers is significantly different from zero. Moreover this effect of evaluation interview still remains significant when one corrects for the selection effect. As stated by prediction 10, evaluated workers have a better knowledge of the rules driving wage setting.

4. Conclusion

In this paper we proposed a theoretical framework based on a Principal-Agent model to analyse the underlying mechanisms of individual evaluation interviews in the case of individual production and of team production. We distinguish an *ex post* evaluation interview that builds a subjective evaluation of employees' effort and an *ex ante* evaluation interview which, in the case of team production works as a coordination device through the fostering of a team spirit. This theoretical analysis allows deriving testable predictions regarding the effect of individual evaluation interviews on productive and cognitive effort, on work overload and on wage setting.

Using a matched employer / employee survey on computerisation and organisational change (COI), we are able to test part of these predictions and to corroborate them. First, evaluation interviews have a positive impact on productive and cognitive effort. Second, evaluation interviews increase effort through two effects: the classical incentive effect and also a selection effect. Third, the selection effect is stronger in the case of individual production compared with the case of team production. Forth evaluated employees earn more than employees in a classical incentive scheme and fifth, evaluated workers have a better knowledge of the rules driving wage setting.

Our theoretical model also suggests a higher propensity to evaluate workers in firms when the production technology is of a super-modular type and an overintensification of work in such a technological context. Our empirical results tend to indicate that collective work is positively correlated with super modular technologies as collective workers are more frequently evaluated and provide a higher level of effort than individual workers. However, evaluation interviews are not associated with work overload. On the contrary, individual workers seem to be protected from work overload when they are evaluated and collective workers do not register a higher exposure to work overload even though they provide higher levels of effort. This could be an indirect evidence of the selection effect already stressed. Evaluated workers produce higher levels of effort, but their personal characteristics or the characteristics of their employers allow them to better cope with it.

References.

- 1) Abowd J. M., Creecy R. H., Kramarz F. (2002): "Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data", *LEHD technical paper*, 2002-06.
- 2) Alchian A., Demsetz H. (1972): "Production, Information Costs, and Economic Organization", *American Economic Review*, Vol.62, pp.777-795.
- 3) Che Y-K., Yoo S-W. (2001): "Optimal Incentives for Teams", *American Economic Review*, Vol. 91, pp.525-541.
- 4) Crifo P., Diaye M.-A., Greenan N. (2004): "Pourquoi les entreprises évaluent-elles individuellement leurs salariés ?" *Economie et Prévision*, N 164-165, Vol.3-4, pp.27-55.
- 5) Diaye M.-A., Greenan N., Urdanivia M. W. (2006): "Subjective evaluation of performance through individual evaluation interview: theory and empirical evidence", CEE Working Paper, N.67, September.
- 6) Engellandt A., Riphahn R.T. (2004): "Incentive Effects of Bonus Payments: Evidence from an International Company", Miméo University of Basel, presented at the First Annual IZA Workshop on Behavioral and Organizational Economics (Bonn, May 30-31 2005).
- 7) Greenan N., Hamon-Cholet S. (2001): "Un dispositif d'enquêtes couplées employeurs/employés sur les changements organisationnels et l'informatisation.", Miméo Centre d'Etudes de l'Emploi.
- 8) Hamilton B. H., Nickerson J. A., Owen H. (2003): "Team Incentives and Workers Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation", *Journal of Political Economy*, Vol.111, pp.465-497.
- 9) Heckman J. J., Ichimura H., Todd P. (1998): "Matching as an Econometric Evaluation Estimator", *Review of Economics Studies*, Vol.65, pp.261-294.
- 10) Heckman J. J., Ichimura H., Todd P. (1997): "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme", *Review of Economics Studies*, Vol. 64, pp.605-654.
- 11) Holmstrom B. (1982): "Moral Hazard in Teams", *Bell Journal of Economics*, Vol. 13, pp.324-340.
- 12) Kandel E., Lazear E.P. (1992): "Peer Pressure and Partnership", *Journal of Political Economy*, Vol.100, pp.801-817.
- 13) Kersley B., Alpin C., Forth J., Bryson A., Bewley H., Dix G., Oxenbridge S. (2005): *Inside the Workplace: First Findings from the 2004 WERS*. Routledge.
- 14) MacLeod B. (2003): "Optimal Contracting with Subjective Evaluation", *American Economic Review*, Vol.93, pp. 216-240.
- 15) Macleod B., Parent D. (1999): "Jobs Characteristics and the Form of Compensation", *Research in Labor Economics*, Vol.18, pp.177-242.
- 16) Osborne M., Rubinstein A. (1994): *A Course in Game Theory*. The MIT Press (3rd printing).
- 17) Prat A. (2002): "Should a team be homogeneous?", *European Economic Review*, Vol.46, pp.1187-1207.
- 18) Prendergast C. (1999): "The Provision of Incentives in Firms", *Journal of Economic Literature*, Vol.37, pp.7-63.
- 19) Rosenbaum P. R., Rubin D. B. (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects", *Biometrika*, Vol.70, pp.41-55.

Appendix 1: Variables constructed from the labour force section of the COI survey

A. Measures of Collective Work.

In the COI survey, a sample of randomly selected employees within interviewed firms (1, 2 or 3 per firm) are asked to describe in detail the way they work at the time when they are being surveyed.

Measure 1: Teamwork

This measure is associated to the following question: "Do you sometimes do your work in group or collectively?" Responses are either "yes" or "no".

Measure 2: Time spent in teamwork

This measure is constructed from the following question asked to employees who declared working in group or collectively: "How much of your working time do you work in group or collectively? Almost all the time, more than a quarter of your time, less than a quarter of your time".

Measure 3: Communication in the firm

This measure is constructed from the following 4 questions: "Apart from your superiors, are there other persons who give you indications on what you have to do? (responses are either "yes" or "no", or it does not apply) (1) "Colleagues you usually work with?" (2) "Other persons or departments in the firm?" "Apart from your subordinates do you give indications to other persons on what they have to do? (responses are either "yes" or "no", or it does not apply) (3) "Colleagues you usually work with?" (4) "Other persons or departments in the firm?" Then a "Low" intensity of communication corresponds to 0 or 1 positive answer among these 4 questions, and a "High" intensity of communication to at least 2 positive answers among the 4 questions.

Measure 4: Support from other workers

The measure is constructed from the following 3 questions: "If you have a temporary excess workload or if you are uneasy with a difficult task are you helped by..." (responses are either "yes" or "no", or it does not apply) (1) "Your superiors?" (2) "Colleagues you usually work with?" (3) "Other persons or departments in the firm?" Then a "Low" level of support from other workers corresponds to 0 or 1 positive answer and a "High" level of support corresponds to at least 2 positive answers among the 3 questions.

Measure 5: Participation into meetings

The measure is constructed from the following question: "How many times a year do you participate into meetings in the context of your work? Then a "Low" participation into meetings corresponds to 0 or only one meeting a year and a "High" participation into meetings corresponds to at least 2 meetings a year.

Table A.1: Five measures of interaction between employees in the work process

	Frequency	Percent
<i>Measure 1: Teamwork</i>		
No (0)	1422	48.97
Yes (1)	1482	51.03
<i>Measure 2: Teamwork intensity</i>		
Less than 1/4 of time (0)	2045	70.42
1/4 of time or more (1)	859	29.58
<i>Measure 3: Communication intensity</i>		
Low (0)	1019	35.09
High (1)	1885	64.91
<i>Measure 4: Level of support</i>		
Low (0)	1537	52.93
High (1)	1367	47.07
<i>Measure 5: Participation into meetings</i>		
Low (0)	1557	53.62
High (1)	1347	46.38

Table A.2: Correlation Coefficients between these five measures of interaction between employees

	measure 1	measure 2	measure 3	measure 4	measure 5
measure 1	1.00000				
measure 2	0.63486	1.00000			
measure 3	0.15159	0.04493	1.00000		
measure 4	0.17993	0.11736	0.14987	1.00000	
measure 5	0.13480	0.04775	0.17168	0.08567	1.00000

B. Measures of effort.

Two dimensions of effort are captured in the COI survey.

The first one describes the **level of productive effort**. It is built from the answers to the two following questions: (1) "do you work more than ordinarily?" (response is either "yes" or "no"); (2) If yes, "do you work more than ordinarily for personal reasons?" (responses is either "yes" or "no").

Employee's effort is then a variable with 3 levels: 0 when the answer to (1) is "no", 1 when the answer to question (1) is "yes" and the answer to question (2) is "no", 2 when both the answers to questions (1) and (2) are "yes".

The second one describes the **level of cognitive effort** or, more precisely the degree of implication into collective knowledge building about the production process. It is built from the answers to the following questions: (1) "in the context of your work, do you make propositions to improve your work station, the production process, the machines...?" (response is either "yes" or "no");

C. Measures of work overload

Two indicators measure work overload.

- *Hurry* is a discrete variable. It is equal to 1 if the employee states that he never has to hurry to do his work. It is equal to 2 if he states that it is the case for less than one quarter of the time. It is equal to 3 if he states that it is the case for one quarter of the time or more. And it is equal to 4 if he states that it he has to hurry almost all the time.
- *Interrupt* is a dummy variable equal to one when the employee states that he often has interrupt one task to carry out another urgent and non anticipated one. It is equal to 0 otherwise.

D. Measure of monetary incentives

Monetary incentives are captured through annualised **net wage** (in euros). It comes from the annual declarations of social data (DADS) which is an administrative file used to compute the tax on wages. It groups all earnings paid in cash or kind between the 1st of January and the 31st of December 1996 less social contributions (social security, pensions, and unemployment benefit).

This compensation includes base wage, all bonuses, taxed allowances and compensations in kind. Bonuses associated with the two French profit sharing regimes (participation and inté ressement) are not included when they are not taxed. However, bonuses connected to participation schemes are generally not taxed when the reverse is true for bonuses connected to inté ressement schemes. It is the length of the period during which bonuses remain unavailable that determines taxation. In the case of participation, when this unavailability period is shortened to three years, the bonuses become partly eligible to taxation. In the case of intéressement, bonuses are partly exonerated from taxes when they are blocked for a while in a company saving scheme.

If we except bonuses connected with participation, compulsory in firms with more than 50 employees, most of the earnings that contribute to an individualization of compensations are

taken into account in our variable. Thus, we may interpret it as an output of the wage policy of the firm.

Lastly, compensations correspond to employment periods that vary from one employee to the other. We have annualised the information we had, taking into account the number of days worked. This does not correct for part time, but only 6% employees in our sample declare working part time.

E. Employee's ability to predict his wage

This variable is built from the answers to the 8 following questions:

"Which of the following elements have a big influence on your wage or on your promotion" (for each element response is either "yes" or "no" or it does not apply): (1) to do a high-quality work?; (2) to carry assignments to the letter?; (3) to be in good terms with the boss (bosses)?; (4) to be in good terms with the colleague(s)?, (5) To take up training courses?, (6) to learn how to use new technologies?; (7) The firm's performances?; (8) other reasons?

Employee's ability to predict his wage is then the ratio of number of "yes" given to these 8 questions to the number of questions where the employee answered "yes" or "no".

Appendix 2: Variables constructed from the Firm section of the COI survey

A. Firms' computerization intensity

This variable is constructed from the question near to the firms: "did/do your firm realize/ realized data transfers by means a computer interface" (responses are either "yes" or "no"): (1) "within the management service?" (2) "between management and production service?" (3) "between management and suppliers, subcontractors?" (4) "between management and client firms?" (5) "between management et social organisms public power?" (6) "between conception services and production?" (7) "between conception and suppliers, subcontractors?" (8) "Within the production services or between manufacture unities?" (9) "between production and suppliers, subcontractors?" (10) "between production and client firms?"

Then intensity 1 corresponds to 0 or 1 "yes"; intensity 2 corresponds to 2 or 3 "yes"; intensity 3 corresponds to 4 or 5 "yes"; and intensity 4 corresponds to 5 or more than 5 "yes". Reference is then intensity 1.

B. Average number of task each type of individual is responsible for (NMT)

This variable is constructed from the question asked to the firms: "In general who is/was authorized in 1997 to...(more than 1 answer possible among Management/Production Worker/Specialist)": (1) adjust installations; (2) perform first level maintenance; (3) allocate tasks to production workers; (4) inspect quality of supplies; (5) inspect quality of production; (6) participate in performance improvements; (7) participate in project teams; (8) stop production in case of an incident; (9) troubleshoot in case of an incident; (10) start production again in case of an incident.

The qualitative variable *NMT* with 4 items is constructed as follows: $NMT \geq 1.7$ (joint responsibility of indirect task) $1.4 \leq NMT < 1.7$ (medium sharing of responsibility) $1 < NMT \leq 1.4$ (low sharing of responsibility) $NMT \leq 1$ (disjoined responsibilities). And reference is $NMT \leq 1$.

Appendix 3: Logistic regression results for the binary outcome "Evaluated/Not evaluated"

A. The case of individual production.

Parameter	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	-2.0329	0.3485	34.0305	<.0001
Socio-demographic characteristics of the employee				
• Gender ^a	-0.2524	0.1519	2.7595	0.0967
• Age ^b :				
15 to 24	-0.1760	0.4762	0.1366	0.7116
25 to 39	-0.0446	0.1963	0.0515	0.8205
40 to 49	0.2130	0.1865	1.3044	0.2534
• Years in the firm ^c :				
1 to 2	0.6301	0.2374	7.0452	0.0079
3 to 6	0.2496	0.1952	1.6355	0.2010
7 to 10	0.3229	0.1756	3.3827	0.0659
• Level of education ^d :				
Vocational training (CAP and BEP)	-0.0175	0.1503	0.0135	0.9074
Second level education (BAC)	-0.0885	0.2581	0.1177	0.7316
Third level education	-0.0496	0.2607	0.0363	0.8489
• Professional type ^e :				
Executives	0.6554	0.3603	3.3089	0.0689
Middle management	0.7697	0.2356	10.6728	0.0011
Clerk	0.3463	0.2461	1.9801	0.1594
Skilled blue collar	-0.0568	0.1670	0.1156	0.7339
General characteristics of the firm				
• Firm size ^f :				
100 to 499	-0.0328	0.1744	0.0355	0.8506
500 to 999	0.0294	0.2264	0.0168	0.8968
1000 and more	0.6202	0.2525	6.0334	0.0140
• Industry sector ^g :				
Mineral products	0.4561	0.3481	1.7165	0.1901
Textile	0.4668	0.3269	2.0391	0.1533
Clothing and leather	-0.2872	0.4127	0.4845	0.4864
Wood and paper	0.1410	0.3348	0.1773	0.6737
Printing, press, publishing	0.2648	0.3902	0.4605	0.4974
Production of propellants and fuels	1.0360	1.4831	0.4880	0.4848
Chemicals, rubber, and plastic products	0.6593	0.2753	5.7342	0.0166
Pharmaceutical, perfumes, and cleaning products	1.7797	0.3673	23.4742	< .0001
Foundry and metal work products	-0.0104	0.2843	0.0013	0.9709
Mechanical engineering	0.1636	0.2718	0.3625	0.5471
Household equipment	0.0894	0.3122	0.0821	0.7745
Electrical and electronic equipment	0.9187	0.4546	4.0840	0.0433
Electrical and electronic components	0.6605	0.2988	4.8870	0.0271
Automobile	0.3523	0.3630	0.9420	0.3318
Shipbuilding, aircraft, and railway	0.6672	0.3829	3.0367	0.0814

^a Reference is "men". ^b Reference is "50 and more". ^c Reference is "11 and more". ^d Reference is "with no degree except CEP or BEPC". ^e Reference is "unskilled blue collar". ^f Reference is "99 and less". ^g Reference is "food industries".

Appendix 3: Logistic regression results for the binary outcome "Evaluated/Not evaluated"

A. The case of individual production.

Parameter	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Firms' computerization intensity ^h				
Intensity 2	0.0295	0.1914	0.0238	0.8773
Intensity 3	0.1274	0.2034	0.3923	0.5311
Intensity 4	0.0664	0.2266	0.0860	0.7694
Obstacles to the organizational changes				
Tensions between the services ⁱ	-0.2390	0.1760	1.8428	0.1746
Tensions with the shareholders ⁱ	0.0501	0.2123	0.0558	0.8133
Difficulties in the relations with the other firms ⁱ	0.0393	0.1899	0.0429	0.8359
Difficulties to school or to reclassify the staff ⁱ	-0.0221	0.1621	0.0186	0.8915
Non executive staff adaptations and establishment problems ⁱ	-0.0908	0.1751	0.2688	0.6041
Executive staff adaptations and establishment problems ⁱ	0.2754	0.1679	2.6886	0.1011
Clashes with the staff (petitions, strikes,...) ⁱ	-0.0970	0.2035	0.2271	0.6337
Use of new organizational devices				
ISO 9001, ISO 9002, EAQF Certification ^j	0.4734	0.1616	8.5854	0.0034
Other certification or total quality management ^j	0.0457	0.1394	0.1073	0.7432
Value analysis, functional analysis, or "AMDEC" method ^j	0.00832	0.1633	0.0026	0.9593
5S method or Total Productive Maintenance (TPM) method ^j	0.2457	0.1768	1.9307	0.1647
Organization in profit centers ^j	0.1212	0.1398	0.7517	0.3859
Formal in-house customer/ supplier contracts ^j	0.0257	0.1402	0.0335	0.8547
System of "Just in time" delivery ^j	0.1323	0.1786	0.5483	0.4590
System of "Just in time" production ^j	-0.0757	0.1793	0.1785	0.6727
Evolution in hierarchical layers between 1994 and 1997 ^k				
1 and more	-0.0725	0.2648	0.0750	0.7842
-1	-0.1617	0.1748	0.8556	0.3550
-2 and less	0.1750	0.2874	0.3707	0.5426
Team work				
• Share of production workers participating in self managed teams ^l :				
10% to less than 50%	-0.1373	0.1965	0.4879	0.4849
50% and more	0.1451	0.2585	0.3149	0.5747

^h Reference is "intensity 1". See A. of appendix 2 for the construction of this variable.

ⁱ The variable is equal to 1 when the firms states that such an obstacle has been either "quite important", or "important", or "very important", and 0 when she states that it has been "unimportant".

^j Response is either "yes" or "no". ^k Reference is "0". ^l Reference is "less than 10%".

Appendix 3: Logistic regression results for the binary outcome "Evaluated/Not evaluated"

A. The case of individual production.

Parameter	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
• Share of production workers participating in problem solving groups ^l :				
10% to less than 50%	0.3005	0.1927	2.4314	0.1189
50% and more	0.5594	0.3928	2.0281	0.1544
• Share of others workers participating in project teams ^l :				
10% to less than 50%	0.1398	0.1958	0.5101	0.4751
50% and more	-0.9496	0.5843	2.6415	0.1041
• Share of others workers participating in self managed teams ^l :				
10% to less than 50%	0.1439	0.2051	0.4921	0.4830
50% and more	-0.1158	0.3478	0.1108	0.7392
• Share of others workers participating in problem solving groups ^l :				
10% to less than 50%	-0.1922	0.2010	0.9143	0.3390
50% and more	-0.0789	0.3770	0.0438	0.8343
• Share of others workers participating in project teams ^l :				
10% to less than 50%	-0.2005	0.1941	1.0673	0.3016
50% and more	0.3645	0.3384	1.1598	0.2815
• Who is/was authorized in 1997 to participate in project teams?:				
Management ^m	-0.2209	0.1608	1.8877	0.1695
Production worker ^m	0.0596	0.1448	0.1694	0.6806
Specialist ^m	-0.2073	0.1627	1.6232	0.2027
• Average number of task each type of individual is responsible for ⁿ :				
1.1 to 1.4	-0.0469	0.1864	0.0635	0.8011
1.5 to 1.7	0.1619	0.2226	0.5292	0.4669
1.8 and more	0.3439	0.2386	2.0764	0.1496

^l Reference is "less than 10%".

^m Response is either "yes" or "no".

ⁿ Reference is " ≤ 1 ". See B. of appendix 2 for the construction of the variable.

Appendix 3: Logistic regression results for the binary outcome "Evaluated/Not evaluated"

B. The case of team production.

Parameter	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	-1.7432	0.3745	21.6633	< .0001
Socio-demographic characteristics of the employee				
• Gender ^a	-0.1403	0.1523	0.8481	0.3571
• Age ^b :				
15 to 24	0.5015	0.4257	1.3880	0.2387
25 to 39	0.2781	0.2295	1.4691	0.2255
40 to 49	0.2781	0.2295	1.4691	0.2255
• Years in the firm ^c :				
1 to 2	-0.2096	0.2443	0.7363	0.3908
3 to 6	0.1818	0.1820	0.9972	0.3180
7 to 10	0.0170	0.1730	0.0097	0.9217
• Level of education ^d :				
Vocational training (CAP and BEP)	0.1829	0.1567	1.3621	0.2432
Second level education (BAC)	0.4481	0.2421	3.4263	0.0642
Third level education	0.5279	0.2645	3.9823	0.0460
• Professional type ^e :				
Executives	0.5416	0.3485	2.4147	0.1202
Middle management	0.1200	0.2284	0.2761	0.5993
Clerk	-0.0121	0.2972	0.0017	0.9675
Skilled blue collar	-0.1469	0.1743	0.7101	0.3994
General characteristics of the firm				
• Firm size ^f :				
100 to 499	0.3510	0.1765	3.9524	0.0468
500 to 999	0.7059	0.2080	11.5179	0.0007
1000 and more	0.1941	0.2422	0.6426	0.4228
• Industry sector ^g :				
Mineral products	-0.2853	0.3202	0.7942	0.3728
Textile	0.3355	0.4096	0.6708	0.4128
Clothing and leather	-0.1220	0.3734	0.1068	0.7439
Wood and paper	-0.4769	0.3661	1.6974	0.1926
Printing, press, publishing	-0.8333	0.4390	3.6032	0.0577
Production of propellants and fuels	1.2745	0.9040	1.9877	0.1586
Chemicals, rubber, and plastic products	0.6759	0.2593	6.7962	0.0091
Pharmaceutical, perfumes, and cleaning products	1.2302	0.3578	11.8237	0.0006
Foundry and metal work products	-0.2956	0.2752	1.1535	0.2828
Mechanical engineering	-0.0338	0.2637	0.0164	0.8980
Household equipment	0.1161	0.3083	0.1418	0.7065
Electrical and electronic equipment	0.3344	0.3679	0.8263	0.3633
Electrical and electronic components	0.1719	0.2950	0.3396	0.5601
Automobile	-0.2314	0.3329	0.4832	0.4870
Shipbuilding, aircraft, and railway	-0.6255	0.3706	2.8479	0.0915

^a Reference is "men". ^b Reference is "50 and more". ^c Reference is "11 and more". ^d Reference is "with no degree except CEP or BEPC". ^e Reference is "unskilled blue collar". ^f Reference is "99 and less". ^g Reference is "food industries".

Appendix 3: Logistic regression results for the binary outcome "Evaluated/Not evaluated"

B. The case of team production.

Parameter	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Firms' computerization intensity ^h				
intensity 2	0.3321	0.1888	3.0940	0.0786
intensity 3	0.4203	0.1960	4.5997	0.0320
intensity 4	0.3323	0.2140	2.4118	0.1204
Obstacles to the organizational changes				
Tensions between the services ⁱ	-0.0540	0.1773	0.0927	0.7608
Tensions with the shareholders ⁱ	-0.2445	0.2312	1.1179	0.2904
Difficulties in the relations with the other firms ⁱ	-0.0976	0.1999	0.2382	0.6255
Difficulties to school or to reclassify the staff ⁱ	0.0598	0.1538	0.1513	0.6973
Non executive staff adaptations and establishment problems ⁱ	0.0411	0.1639	0.0629	0.8020
Executive staff adaptations and establishment problems ⁱ	0.1569	0.1589	0.9750	0.3234
Clashes with the staff (petitions, strikes,...) ⁱ	-0.1195	0.1930	0.3833	0.5358
Use of new organizational devices				
ISO 9001, ISO 9002, EAQF Certification ^j	0.4089	0.1604	6.4964	0.0108
Other certification or total quality management ^j	0.1545	0.1389	1.2379	0.2659
Value analysis, functional analysis, or "AMDEC" method ^j	-0.0932	0.1582	0.3470	0.5558
5S method or Total Productive Maintenance (TPM) method ^j	0.4285	0.1631	6.8979	0.0086
Organization in profit centers ^j	0.1763	0.1351	1.7015	0.1921
Formal in-house customer/ supplier contracts ^j	0.1045	0.1380	0.5728	0.4492
System of "Just in time" delivery ^j	-0.3277	0.1778	3.3980	0.0653
System of "Just in time" production ^j	0.1577	0.1781	0.7846	0.3757
Evolution in hierarchical layers between 1994 and 1997 ^k				
1 and more	-0.0279	0.2610	0.0114	0.9150
-1	-0.0204	0.1629	0.0156	0.9005
-2 and less	-0.3818	0.2959	1.6652	0.1969
Team work				
• Share of production workers participating in self managed teams ^l :				
10% to less than 50%	0.0247	0.1780	0.0193	0.8895
50% and more	0.0651	0.2613	0.0620	0.8033

^h Reference is "intensity 1". See A. of appendix 2 for the construction of this variable.

ⁱ The variable is equal to 1 when the firms states that such an obstacle has been either "quite important", or "important", or "very important", and 0 when she states that it has been "unimportant".

^j Response is either "yes" or "no".

^k Reference is "0".

^l Reference is "less than 10%".

Appendix 3: Logistic regression results for the binary outcome "Evaluated/Not evaluated"

B. The case of team production.

Parameter	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
• Share of production workers participating in problem solving groups ^l :				
10% to less than 50%	0.4672	0.1863	6.2869	0.0122
50% and more	0.8599	0.3590	5.7362	0.0166
• Share of production workers participating in project teams ^l :				
10% to less than 50%	-0.0944	0.1838	0.2640	0.6074
50% and more	-0.0239	0.4492	0.0028	0.9575
• Share of others workers participating in self managed teams ^l :				
10% to less than 50%	-0.3616	0.1983	3.3244	0.0683
50% and more	-0.2979	0.3314	0.8082	0.3687
• Share of others workers participating in problem solving groups ^l :				
10% to less than 50%	0.1016	0.2065	0.2419	0.6229
50% and more	0.2644	0.3815	0.4801	0.4884
• Share of others workers participating in project teams ^l :				
10% to less than 50%	0.1084	0.1851	0.3428	0.5582
50% and more	0.0786	0.3393	0.0537	0.8168
• Who is/was authorized in 1997 to participate in project teams?				
Management ^m	-0.3374	0.1701	3.9368	0.0472
Production worker ^m	-0.1134	0.1475	0.5912	0.4420
Specialist ^m	0.1610	0.1591	1.0239	0.3116
• Average number of task each type of individual is responsible for ⁿ :				
1.1 to 1.4	0.0988	0.1917	0.2656	0.6063
1.5 to 1.7	0.1538	0.2221	0.4796	0.4886
1.8 and more	-0.1319	0.2414	0.2984	0.5849

^l Reference is "less than 10%".

^m Response is either "yes" or "no".

ⁿ Reference is " ≤ 1 ". See B. of appendix 2 for the construction of the variable.