



HAL
open science

Feedback and incentives: Experimental evidence

Tor Eriksson, Anders Poulsen, Marie Claire Villeval

► **To cite this version:**

Tor Eriksson, Anders Poulsen, Marie Claire Villeval. Feedback and incentives: Experimental evidence. Labour Economics, 2009, 16 (6), pp. 679-688. halshs-00451557

HAL Id: halshs-00451557

<https://shs.hal.science/halshs-00451557v1>

Submitted on 17 Feb 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

FEEDBACK AND INCENTIVES: EXPERIMENTAL EVIDENCE

Tor Eriksson^a

^a Department of Economics, Aarhus School of Business, Aarhus University: Frichshuset, Hermodsvej 22, 8230 Åbyhøj, Denmark.
E-mail: tor@asb.dk

Anders Poulsen^b

^b School of Economics, University of East Anglia: Norwich NR4 7TJ, United Kingdom.
E-mail: a.poulsen@uea.ac.uk.

Marie Claire Villeval^c

^c Corresponding author. University of Lyon, CNRS, GATE: 93, Chemin des Mouilles 69130 Ecully, France. IZA, Bonn, Germany. Tel. : +33 472 86 60 79. Fax : +33 2472 86 60 90.
E-mail: villeval@gate.cnrs.fr.

JEL Classification : C70, J16, J24, M52, J33, J31, C91

Keywords: Tournament, piece rate, peer effects, feedback, evaluation, experiment.

This version: 28 July 2009

The authors are grateful to participants at the Conference of the European Association of Labour Economists in Amsterdam, the NBER 2008 Summer Institute, National Chung Hsing University (Taichung), the 3rd Nordic Conference on Behavioral and Experimental Economics in Copenhagen, the Workshop on the Economics of the Workplace at the Tinbergen Institute, the 2nd European Workshop on Experimental and Behavioral Economics in Lyon, the ANR workshop on Incentives and Motivation in Paris, the Conference on Tournaments, Contests and Relative Performance Evaluation in Raleigh, for their comments. We thank two anonymous referees whose suggestions significantly improved the paper. Financial support from the EMIR program of the French National Agency for Research (ANR) and from the Danish Social Science Research Council is gratefully acknowledged.

Abstract

This paper experimentally investigates the impact of different pay schemes and relative performance feedback policies on employee effort. We explore three feedback rules: No feedback on relative performance, feedback given halfway through the production period, and continuously updated feedback. We use two pay schemes, a piece rate and a tournament. We find that overall feedback does not improve performance. In contrast to the piece-rate pay scheme there is in tournaments some evidence of positive peer effects in tournaments since the underdogs almost never quit the competition even when lagging significantly behind, and front runners do not slack off. But in both pay schemes relative performance feedback reduces the quality of the low performers' work; we refer to this as a "negative quality peer effect".

1. INTRODUCTION

An important aspect of the strategy of most organizations is the provision of incentives to the employees to meet the organization's objectives. Typically this implies tying pay to performance (see Prendergast, 1999). In order to reward employees for their effort, firms spend considerable resources on performance evaluations. In many cases, evaluation consists of comparing actual performance to a predefined individual target. Another frequently used format is relative performance evaluation. Relative performance evaluation may motivate employees to work harder. But it may also be demoralizing and create an excessively competitive workplace, which may hinder overall performance; see Lazear (1989). Determining the overall impact of relative performance evaluation is crucial for companies. Economic research on relative performance evaluation has mainly focused on the comparison of final performances between competitors, like in tournament theory, and on quantitative and subjective performance ratings (Lazear and Gibbs, 2009). In contrast, what happens during a competition and the impact of feedback frequency on effort have so far received little attention in the literature. The aim of this study is to add to this literature.

We address two issues. How intensively should the organization provide its employees with feedback on their relative performance, when the organization's objective is to maximize the employees' output? How does the answer to this question depend on the pay scheme in use? We examine these questions experimentally by studying individuals' performance under a piece rate and a tournament pay scheme, and under three feedback rules, namely no feedback, feedback given half-way during the production period, and continuous feedback on relative performance. Which combination of pay scheme and feedback rule yields the highest output and profit?

Pay and relative performance feedback are crucial aspects of organizational design. First, feedback on fellow employees' performance is likely to matter when the organization uses

pay schemes like tournaments. In such settings the organization must decide whether to inform employees about their relative standing during the competition, or to let them remain ignorant until the end. On the one hand, an agent who is informed he is an underdog may be discouraged and even quit the job. And a front runner who learns that he is well ahead of the other contestants may think that he can afford to slack off. In both cases, performance declines. On the other hand, it is possible that an informed underdog works harder than when uninformed, in order to preserve pride and avoid shame, and that an informed front runner becomes more enthusiastic and increases effort. In this case feedback leads to better performance. Only a few theoretical models have analyzed these issues and they usually conclude that it is best not to give feedback on relative performance (see e.g., Lizzeri et al., 2002, on the impact of feedback on the agent's own performance on effort, and Ederer, 2004, on the impact of feedback on relative performance).¹ The empirical literature on the magnitude of these effects is scant and the current paper is one of the first attempts to provide some empirical evidence on the impact of feedback on performance in the context of performance-pay.²

Second, feedback on relative performance has been found to influence behavior even when conventional economic analysis says it should not matter. Falk and Ichino (2006) and Mas and Moretti (2009) show that allowing people to observe each other's performance has a positive influence on an employee's performance although her payment is independent of the others' performance. This is attributed to positive peer effects (see Kandel and Lazear, 1992), according to which employees adjust their effort to the co-worker's effort, due to non-

¹ See Goltsman and Mukherjee (2008) on the differentiation of announcements according to the agents' difference in ability, Yildirim (2005) on optimal interim revelation of ability by the agents themselves, Ertac (2005) on the information derived by the agents from the feedback given about the other agents, and Gershkov and Perry (2009) on the optimal aggregation rule between the midterm and final reviews. Another strand of the literature focuses on biased feedback. An early treatment of whether an organization can benefit for distorting information is Gibbs (1991); see Gürtler and Harbring (2007) for experimental evidence.

² The management literature, see e.g., Milkovich and Newman (1996), has a clear message: revealing true performance information to the employee is bad. The standard explanation, based on psychology and politics, is that the information creates animosity among employees and hurts morale.

monetary reasons such as social pressure, fear of informal sanctions, and shame. In this paper we use a similar definition of peer effects. Social psychologists have documented many non-monetary effects of relative performance feedback; see for example Kluger and Denisi (1996). Economists have recently started to formally model how peer effects interact with monetary incentives (see e.g. Kräkel, 2007).

What is the role of the pay scheme and relative performance feedback for performance and profit? If providing feedback on relative performance makes a difference, how is this decomposed into the separate impacts on underdogs and on front runners? Are the positive peer effects that have been documented for performance-independent pay schemes (Falk and Ichino, 2006 and Mas and Moretti, 2009) also present in performance-pay schemes or are they crowded out by the stronger monetary incentives? Our paper sheds empirical light on these issues. Since it is difficult to get suitable data from real organizations we conduct a laboratory experiment.³ We find that on the whole, feedback on relative performance, regardless of the performance-pay scheme used, does not improve performance. Neither discrete nor continuous feedback on relative performance generates positive peer effects in the piece-rate pay scheme. Additional experimental sessions indicate, however, that this absence of peer effects in the piece rate pay scheme may be due to subjects exerting maximum effort given their ability, in which case there is little room for positive effects to manifest themselves. In contrast, we find evidence of positive peer effects in the tournaments: front runners do not slack off and underdogs almost never quit the competition, even when lagging significantly behind. These effects are however not sufficient to improve overall performance when compared to an environment with no feedback. In addition, in both pay schemes there is what we call a negative ‘quality peer effect’: feedback on relative performance has a negative effect on scores, not because it causes the less able performers to

³ Notable exceptions are Bandiera, Barankay, and Rasul (2005), Lazear (1992), and Mas and Moretti (2009).

work less hard but because they make more mistakes. We believe our experiment is the first to provide evidence on how feedback can affect not only the quantity but also the quality of agents' effort.

The rest of the paper is organized as follows. Section 2 describes the literature on feedback. Section 3 describes the experimental design and formulates some hypotheses. The data are analyzed in Section 4 and Section 5 discusses our results and concludes.

2. RELATED LITERATURE

Falk and Ichino (2006) and Mas and Moretti (2009) examine the impact of feedback on relative performance on effort in settings with a flat wage pay scheme. Both studies find significant positive peer effects. In the first study subjects stuff envelopes in exchange for a fixed payment. Subjects either work alone or work in pairs where each subject can see the other's output. Falk and Ichino observe that output is higher in the latter case and interpret this as evidence of positive peer effects. Their data are consistent with a model where workers experience disutility when they fall behind their fellow workers.

Mas and Moretti examine peer effects in settings where super market cashiers are affected by other workers' performance through externalities. Each cashier works and is paid independently of each other, but has an incentive to free ride by unloading the workload on to the other workers. The data show that more productive workers induce an increase in productivity of fellow employees, but only significantly so when the workers can observe each other's output. This is consistent with peer effects being social pressure (including mutual monitoring or informal sanctions).

The two papers described above use a flat wage pay scheme. We empirically examine whether or not peer effects are present when the organization uses performance-pay. The few studies that have considered the influence of competitors' ability on own effort have produced mixed evidence; see Haraguchi and Waddell, 2007, Guryan et al. 2009, and Ludwig and

Lünser, 2009. Bandiera, Barankay, and Rasul (2005) find that when a relative pay scheme is used, a worker's productivity is affected by her co-workers' productivity as the negative externality is internalized, but only when mutual monitoring is possible.

Ederer and Fehr (2007) set up a two-period tournament experiment where a principal can at no cost give potentially incorrect feedback on the relative performance of two workers. Feedback is provided at the end of period 1, after which workers decide on their period 2 effort. In a second treatment feedback is always correct and in a third treatment no feedback is given. According to their theoretical analysis, if marginal cost is linear (convex) [concave] then period 2 effort under accurate feedback is the same as (lower than) [higher than] under no feedback. Like Ederer and Fehr we compare tournaments with no feedback and with feedback, but in our experiment feedback is always correct. In addition, we study the case where feedback is given continuously. This allows us to identify the potential influence of the frequency of feedback on behavior. Moreover, whereas the subjects in Ederer and Fehr's experiment choose abstract effort levels our subjects work on a real task. While we thereby lose experimental control over the cost of effort, it allows us to observe the dynamics of effort over time in the Continuous Feedback treatment, where workers are constantly kept updated on relative performance. Finally, since our real task allows us to measure how many errors the subjects make we can investigate the extent to which feedback affects performance for given effort, which in turn provides us with a simple measure of the quality of effort.

Some papers have investigated tournaments where one contestant is exogenously favored over another player (Schotter and Weigelt, 1992, Kräkel, 2007). They find that the disadvantaged contestant does not frequently drop out and chooses a larger effort than predicted by theory. In contrast, Müller and Schotter (2007) find that low-ability workers drop out when promotion opportunities are limited. We do not exogenously impose differences between contestants. Instead, such differences emerge endogenously as players

exert effort and observe their relative positions. We feel that using such a dynamic experiment provides a better framework for studying quitting decisions.

Fershtman and Gneezy (2005) study quitting behavior by conducting an experiment with students running athletics races. The monetary stakes were manipulated and the tournaments were either direct (runners ran side by side) or indirect (subjects ran separately). The latter distinction is similar to our No Feedback and Continuous Feedback treatments. It is found that it does not matter for performance whether the tournament is direct or indirect, but there is less quitting in the indirect tournament. Dropping out occurs only when incentives are strong. The low drop out rate is attributed to shame due to the presence of an audience. Our paper studies quitting in settings where subjects receive a feedback on relative performance but there is no audience.

3. EXPERIMENTAL DESIGN

3.1. Procedures

Our experiment uses a real task that consists of adding sets of four randomly generated two-digit numbers for 20 minutes. All the subjects in all sessions receive the same numbers in the same order. This task is similar to the one used in Niederle and Vesterlund (2007). We chose the task because it is simple, we expected it to be gender-neutral⁴, it requires effort (concentration) but not learning, the outcome is easy to measure, and it provides us with a simple measure of the quality of effort (wrong versus correct answers). The subjects see the numbers on the computer screen and enter the answer using the keyboard. Use of a pen, scratch paper, or calculator is not allowed. Once an answer has been submitted, a new set of numbers appears on the screen. Each subject can work at his own pace. The screen displays the number of correct answers and the remaining time available.

If in a tournament a competitor feels he is certain to lose he may decide that it is not

⁴ Gender neutrality is confirmed by our experiment where scores do not differ across gender ($p=0.363$).

worthwhile to continue performing the task. But simply sitting still in the lab may not be an attractive alternative either. We made a newspaper available on each desk and informed the subjects that they were free to read this newspaper, or any book that they had brought with them. A subject could read for as much and as long as they wished, and they could resume work if they so desired.

Pay schemes

We randomly matched subjects in pairs. In the Piece Rate pay scheme a subject's earnings only depend on the subject's own score (number of correct answers). Each correct answer pays 1 point (10 points=€1). There is no penalty for wrong answers. In the Tournament pay scheme a subject's earnings depend on the relative performance. The subject with the highest score receives a fixed winner's prize of €17 and the other subject receives €0. In case of ties the winner is randomly selected and the other subject receives nothing.

Feedback rules

Subjects always know their own current score but the information they get about the other subject's score depends on the feedback rule. Under the No Feedback rule a subject never observes the other subject's performance. Under the Discrete Feedback rule each subject is informed about the other subject's score halfway during the production period, that is, when 10 minutes have elapsed; after that no further feedback is given before the end of the game. According to the Continuous Feedback rule each subject is continuously informed about the other subject's score. Regardless of the feedback rule a subject observes the other's score after the 20 minutes have elapsed, except in the Piece Rate No - Feedback treatment, and in the Tournament pay scheme he is also told if he won or lost. All the above is public information. Appendix 1 contains a screen shot. All in all, we therefore have $2 \times 3 = 6$ treatments, of the form (x,z) , where $x=PR$ (Piece Rate), T (Tournament), and $z=NF$ (No Feedback), DF (Discrete Feedback), CF (Continuous Feedback).

Choosing Piece Rate and Tournament parameters

To enable profitability comparisons across treatments we chose the parameters of the treatments (per unit payment and prizes) such that the expected wage cost of an employee is the same under both the Piece Rate and Tournament pay schemes. To do this we first fixed the piece rate, b , and ran the Piece Rate - No Feedback treatment. The average observed output from this treatment, denoted \bar{y} , determines an average wage, $b\bar{y}$, which was approximately equal to €8.5. We used the same average wage cost in the tournaments, by setting the loser's prize equal to zero and setting the winner's prize equal to $2 \times €8.5 = €17$. This equalizes the expected wage cost per worker-pair in the Tournament and in the Piece Rate pay schemes, assuming that feedback has no impact on output in the Piece Rate pay scheme. This, in turn, implies that profitability differences exactly reflect differences in scores.

Experimental procedure

The experiment was conducted at the GATE laboratory in Lyon, France, using the REGATE software (Zeiliger, 2000). 208 undergraduate students from local business and engineering schools (101 males and 107 females) were recruited by means of the ORSEE software (Greiner, 2004). Eight sessions were organized. Each subject participated in a single session only. Table 1 summarizes the treatments.

(Table 1 about here)

Upon arrival the subjects were randomly assigned to a computer. The instructions, describing the nature of the task, the money payoffs and the nature of feedback were distributed and read aloud (they are available from the authors upon request). The subjects completed a questionnaire in order to check their understanding. They practiced the task for three minutes without payment so they could become familiar with the screen display and the task. Next, each subject was randomly matched with another subject and performed the task for a maximum of 20 minutes, with or without feedback on the co-worker's score, depending

on the feedback rule. A session lasted on average 75 minutes. Subjects received a show-up fee of €7. On average the subjects earned € 15.41 (min=€7; max=€24). Subjects were paid in cash in a separate room.

3.2. Predictions

In this section we formulate some hypotheses about the behavior we expect to see in our treatments. Our objective is not to set up and test a theoretical model. Any such attempt would ultimately be unsuccessful since in our real effort experiment we do not have full experimental control over the parameters which according to theory are important, such as the cost of effort. Our experiment empirically elucidates how individuals react to interim feedback in dynamic environments.

Peer effects

By peer effects we understand a change in people's submissions due to non-monetary reasons triggered from observing other agents' scores. People may compare their score with reference agents, and it can be subjectively costly to be out of line with them (Kandel and Lazear, 1992; Kräkel, 2007). If peer effects are positive then lagging behind a reference agent is unpleasant, and in order to avoid this feeling the agent increases effort in order to catch up with the other agent. Similarly, outperforming others is pleasant, and the agent seeks to increase his distance with the other agent. If in contrast peer effects are negative, lagging behind leads to discouragement and the agent reduces effort, or even quits although it may in terms of maximization of expected money earnings be rational to continue exerting effort. As Falk and Ichino (2006) we think of peer effects as affecting the marginal cost of effort. Peer effects are positive if a person's marginal cost of effort falls when the other agent increases his score. They are negative if her marginal cost of effort increases when the other agent improves his score. *Ceteris paribus*, positive peer effects lead the agent to increase her submissions whereas negative peer effects lead her to reduce them.

Positive peer effects, and indeed any predicted effort response to changes in the exogenous parameters, can only occur if agents are not already exerting maximum effort. This is an important qualifier, because in any real effort experiment an interior solution cannot simply be assumed. We allow our subjects to freely decide on their submissions and an alternative activity exists (reading newspapers or doing nothing), but agents may have exerted maximum effort (a corner solution), constrained by their ability level. We address this issue in Section 4.1 below.⁵

Since the subjects belong to the same pool of students and are randomly allocated to the various treatments, the ability distributions across treatments are assumed to be the same.⁶ Consequently, observed differences in submissions across treatments will reflect differences in effort, not in ability.

Keeping these observations in mind, we proxy effort by the number of submitted answers. If an answer is false no output is produced: only correct answers generate scores.⁷ In other words, $\text{effort (submitted answers)} = \text{score (correct answers)} + \text{mistakes (incorrect answers)}$. An increase in submissions will not be accompanied by a similar increase in scores if the number of mistakes also increases. In our experiment there is therefore an additional potential peer effect, which affects the ‘quality’ of effort, namely the extent to which a given effort is translated into scores. In the rest of the paper, ‘*peer effect*’ refers to the peer effect operating on the quantity of submissions, and we refer to the one operating on quality as a ‘*quality peer effect*’. The quality peer effect is defined as being negative when an increase in the other agent’s score leads the agent to make more mistakes. A positive quality peer effect means

⁵ On the relative roles of effort and ability, see Arvey (1972), and the psychology literature on motivation, and expectancy theory in particular, cited therein.

⁶ For the same reason, we do not expect there to be systematic differences in the subjects’ psychological traits across treatments.

⁷ We assume here that mistakes are not due to a lack of effort since if they are not willing to put much effort in the task, the subjects can simply work more slowly or even stop working, which both should reduce the number of mistakes. In contrast, putting more effort in the task, i.e., trying to submit more answers within the time span, may result in subjects making more mistakes because this requires more concentration.

that the agent makes relatively fewer mistakes.⁸ We expect the quality peer effect to be predominantly negative.

The Piece Rate pay scheme

There is no strategic interaction between the agents under the Piece Rate pay scheme, so a prediction based on standard assumptions is that each agent's behavior is the same regardless of the feedback policy.

Prediction 1: Suppose there are no peer effects. Then the average number of submissions and scores in the PR-NF, PR-DF, and PR-CF treatments are the same.

Consider next peer effects. Positive (negative) peer effects in the Piece Rate pay scheme means that a person's optimal choice of submissions is increasing (decreasing) in the other person's score. Of course it is not possible to give clean theoretical results without knowing the shape of the cost of effort function. We instead predict that in the PR-DF treatment there will be sufficient evidence for positive peer effects if subjects' period 2 submissions are observed to be increasing in the other subjects' period 1 score, provided a limit is not imposed by their ability level.⁹ The presence of positive (negative) peer effects should increase (decrease) the overall number of submissions relative to the PR-NF treatment. To sum up:

Prediction 2: Suppose there are positive peer effects. Then in the PR-DF treatment subjects' period 2 submissions and scores will be increasing in the other subjects' period 1 scores.

By extension, we predict that peer effects, if they exist, should have a stronger impact in the PR-CF treatment since subjects receive more frequent feedback on their relative position than

⁸ As for the 'normal' peer effect, it seems natural to define the quality peer effect as the impact on behavior that arises from changes in the marginal cost of some variable that is related to the quality of effort (the likelihood with which the person submits an incorrect answer). But, of course, in order to do this we would need first to construct a model where the agent also decides on the quality of effort, and where the agents' payoff function contains a cost term such as the "cost of concentration", or the "cost of being careful". It would be interesting to construct such a model since it would allow us to study the nature of the agents' quantity-quality tradeoff, but this is beyond the scope of the current paper.

⁹ This follows from recalling that positive peer effects means that an increase in the other person's score lowers the agent's marginal cost of effort and hence raises the agent's optimal level of submissions. The agent's period 2 best reply level of submissions will therefore be increasing in the other agent's period 1 score.

in the PR-DF treatment.

As remarked above, while the pattern described in Prediction 2 is sufficient to conclude that positive peer effects exist, it is not a necessary condition due to the possibility of a corner solution.

The Tournament pay scheme

We shall use the following terminology for the Tournament with Discrete Feedback: agent i is the 'front runner' if the subject's period 1 output was the highest, and agent j is the 'underdog' if his period 1 output was the lowest. Ederer and Fehr (2007) compare a tournament with no feedback with one where agents receive feedback on relative performance halfway through the experiment. Thus, their set-up matches our experimental No Feedback and Discrete Feedback treatments, except for the fact that in our experiment we do not introduce stochastics since this could make the task less transparent for the subjects. Nevertheless, we use the result in Ederer and Fehr (2007) as a benchmark for our own hypotheses. They obtain the following result: In the T-DF treatment and in the absence of peer effects both the front runner and the underdog's period 2 submissions and scores are decreasing in the period 1 score gap.

Consider next the impact of peer effects. In the T-DF or T-CF treatments, when an underdog learns he is lagging behind the front runner, positive peer effects motivate him to increase his number of submissions. This goes against the negative impact on submissions due to the downward revision of the expected monetary earnings. Also, positive peer effects induce the front runner who is aware that he is ahead to increase or at least maintain his number of submissions compared to a front runner not influenced by positive peer effects.¹⁰ This implies that in the T-DF treatment, for a given period 1 output, the period 2 outputs of

¹⁰ Note that we cannot as easily develop a strategy for identifying negative peer effects. These are predicted to lead the underdog to reduce submissions, which is the same that a hypothesis based on no peer effects predicts.

underdog and front runners influenced by positive peer effects exceed the outputs of underdogs and front runners influenced only by monetary incentives. Moreover, under an additional assumption on positive peer effects, which will be maintained below, that the marginal cost of period 2 effort is falling in the period 1 output gap, positive peer effects imply that the period 2 outputs of underdogs as well as front runners is increasing in the period 1 output gap.¹¹ Of course an increase in the number of submissions may not translate into an increased score on a one-to-one basis if it is accompanied by a negative quality peer effect (some effort is then lost through mistakes). To sum up:

Prediction 3A: Suppose there are no peer effects. Then, in the Tournament - Discrete Feedback treatment the period 2 submissions and scores of underdogs and frontrunners decrease in the period 1 output gap. In the Tournament - Continuous Feedback treatment, this overall negative effect should be stronger since it can manifest itself earlier.

Prediction 3B: Suppose there are positive peer effects. Then, in the Tournament-Discrete Feedback treatment the period 2 number of submissions and scores of underdogs as well as front runners increases in the period 1 score gap.

A weaker version of Prediction 3B is that the period 2 submissions are not significantly affected by the interim gap in scores: both the underdogs and front runners keep working at the same pace whatever the score gap. The same behavioral prediction holds for the Continuous Feedback treatment. Finally, with positive peer effects dropping out should be less frequent than in the absence of peer effects, even though the gap in scores is large enough to make the outcome of the competition highly predictable.

Quality peer effects

In both pay schemes it is possible that a positive peer effect is (partially) outweighed by a

¹¹ We thank a referee for drawing our attention to the importance of this assumption. Moreover, as pointed out by the referee, this assumption may not be satisfied if closer period 1 outputs (a “closer race”) leads to more intense social comparisons.

negative quality peer effect. For example, an agent who lags behind and who is driven by comparisons to work harder can at the same time become stressed and anxious, and begin to make more mistakes, in which case some of the extra effort is wasted.¹² In order to detect a quality peer effect in the data we must distinguish it from the peer effect operating on submissions. We propose the following test: Suppose agents' submissions, e , are the same in any two treatments but the proportion of mistakes, m , in the treatment with more intense feedback is larger at the expense of the proportion of scores, s (i.e., m/e is larger and so s/e is smaller). This would provide evidence that increasing feedback on relative performance generates a negative quality peer effect. This gives us the following prediction:

Prediction 4: In the presence of negative quality peer effects, the proportion of mistakes made by the agents who are lagging behind in both pay schemes is the highest in the Continuous Feedback treatments and the lowest in the No-Feedback treatments.

4. RESULTS

We first study the impact of the pay scheme and feedback rule on aggregate performance and on the within-pair differences in performance. We then investigate the presence of peer effects. Finally, we consider quitting decisions.

4.1. Feedback, incentives, and aggregate performance

Table 2 shows average scores (number of correct answers), average mistakes (number of incorrect answers), and average submissions (the sum of scores and mistakes), and their standard deviation.

(Table 2 about here)

We first consider the Piece Rate data. As compared with the No Feedback treatment, providing the subjects with a discrete feedback does not affect the average score ($p=0.970$).

¹²An extreme example of this is the so-called 'choking under pressure' phenomenon (Baumeister, 1984), where an agent's performance deteriorates when the importance of the task is accentuated.

The score is somewhat lower in the Continuous Feedback treatment, but not significantly so ($p=0.558$).¹³ On average, subjects solve 84 problems in PR-NF and in PR-DF, while they only solve 81.5 problems in PR-CF. The difference in the average number of submissions is significantly different neither between the No Feedback rule and the Discrete Feedback rule ($p=0.939$) nor between the No Feedback and the Continuous Feedback rule ($p=0.858$).

In the Tournament pay scheme we find no significant differences in average score across the three feedback policies. The average score is smaller in the Continuous Feedback treatment (76.33) than in the No Feedback treatment (80.30), but the difference is not significant ($p=0.190$). The difference between the average score in the Discrete Feedback treatment (75.93) and the No Feedback treatment is also not significant ($p=0.230$). The difference in the average number of submissions is not significantly different either between T-NF and T-DF ($p=0.462$) or between T-NF and T-CF ($p=0.564$).

Result 1: Giving feedback on the scores does not significantly influence the average score or submissions in neither the Piece Rate nor the Tournament pay scheme.

While this result lends support to Prediction 1 against Prediction 2, we must be careful in interpreting it. As already discussed, Prediction 2 assumes that the subjects can increase their effort after receiving a feedback, i.e. they have not reached a limit imposed by their ability. To investigate the plausibility of this assumption, we ran an additional session for the Piece-Rate No Feedback treatment where we doubled the piece-rate.¹⁴ In this high incentive session, with 16 subjects, the average score is 85.56 and the average number of submissions is 96.89, which can be compared with 84 and 94.86, respectively in the initial data. The subjects work harder in the high incentive session, but the difference is not statistically significant ($p>0.10$). This indicates that in both cases subjects sought to exert maximum effort, implying

¹³ In the non-parametric statistics reported here, each subject in the PR-NF and T-NF treatments gives one independent observation, whereas each pair gives one independent observation in all the other treatments. Unless explicitly mentioned otherwise, these tests are Mann-Whitney rank-sum tests. The p -values for the non parametric tests are two-sided.

¹⁴ We thank an anonymous referee for making this suggestion.

that social comparisons, if they exist, cannot increase performance very much. In other words, our data do not allow us to reject the existence of social comparisons in the Piece Rate pay scheme since positive peer effects, if they exist, may not be able to significantly influence the number of submissions.

Are there, for a given feedback rule, significant score differences between the Piece Rate and the Tournament pay schemes? Table 2 shows that for any feedback rule, average scores are lower in the Tournament than in the Piece Rate treatment. For the No (Discrete) [Continuous] Feedback rule, the average scores is decreased by 4.4 % (9.6 %) [6.3 %], but these differences are not significant (NF: $p=0.429$, DF: $p=0.285$, CF: $p=0.189$). The same applies to the average number of submissions (NF: $p=0.188$, DF: $p=0.256$, CF: $p=0.145$). A Kolmogorov-Smirnov test also rejects the presence of significant differences in the distribution of scores between the two pay schemes in the No Feedback ($p=0.271$), the Discrete Feedback ($p=0.398$), and the Continuous Feedback rules ($p=0.219$).¹⁵

We next consider if there is an impact of feedback on the quality of effort. For the Piece Rate pay scheme the number of mistakes under Continuous Feedback (13.30) is significantly higher than under No Feedback (10.86) ($p=0.049$). Providing discrete feedback does not significantly raise the number of mistakes ($p=0.429$). The percentage of mistakes out of total submissions increases from 11.45% in PR-NF to 12.59 % in PR-DF, and to 14.03% in PR-CF.

In the Tournament the number of mistakes under Continuous Feedback (12.17) is significantly higher than under No Feedback (9.59) ($p=0.005$). Providing feedback halfway has no significant influence ($p=0.150$). The percentage of mistakes in the number of submissions in T-NF, T-DF, and T-CF are 10.67%, 13.01%, and 13.75%, respectively. This

¹⁵ It should, however, be mentioned that if we pool together both feedback rules with interim feedback, we find significant differences between scores in the Piece Rate and the Tournament pay schemes: the average score is lower in the tournament (76.18) than in the Piece Rate pay scheme (82.73) ($p=0.086$). The same conclusions follow when we focus on the number of submissions that is lower in the tournament (88.05) than in the Piece Rate pay scheme (95.43) ($p=0.072$).

yields Result 2, which supports Prediction 4.

Result 2: In both the Piece Rate and the Tournament pay schemes the quality of effort is significantly lower when subjects receive continuous feedback on their relative score than when they receive no feedback.

Note that the number of mistakes is higher in the Piece Rate than in the Tournament pay scheme, but the difference is not significant, neither in the No Feedback treatment ($p=0.367$), nor in the Discrete Feedback ($p=0.930$) nor in the Continuous Feedback ($p=0.761$). Overall, regardless of whether pay depends on absolute or relative score, an ongoing comparison with co-participants' performance does not affect the quantity of effort but reduces its quality.

4.2. Feedback, Incentives, and Within-Pair Performance

We now consider the performance of front runners and underdogs. In what follows, when we report on the final scores, the front runner (underdog) can alternatively be thought of as the winner (loser). Table 3 shows the average scores of the front runner and the underdog in each pair, the absolute average within-pair score difference, the average number of mistakes and total number of submissions made by the front runner and the underdog, by pay scheme and feedback rule.

(Table 3 about here)

Table 3 shows that for each pay scheme the average within-pair score difference is larger when the subjects are continuously informed about their relative score compared to the two other feedback rules. In the Piece Rate pay scheme with No (Discrete) [Continuous] Feedback, the average score differences are 20.28 (22.47) [27.00], respectively. However, the PR-NF treatment does not differ significantly from PR-DF ($p=0.827$), nor from PR-CF ($p=0.336$). In Tournaments the difference between the average score gaps of the Continuous Feedback and No Feedback treatments (30.17 and 14.68) is highly significant ($p=0.003$), while the difference between the Discrete Feedback and No Feedback treatments (22.00 and 14.68) is not significant ($p=0.269$).

What gives rise to the increased within-pair score differences when feedback is continuous? Table 3 shows that for each pay scheme the average score of the front runners is roughly the same across feedback policies ($p=0.868$ between PR-NF and PR-CF, $p=0.628$ between T-NF and T-CF). It thus follows that the increasing average score difference in the intensity of feedback is due to the underdogs reducing their scores.¹⁰

Result 3: In the Tournament pay scheme the within-pair score difference is significantly higher ($p=0.003$), and the score of the underdogs is significantly lower ($p=0.020$) under the Continuous than under the No Feedback rule. In the Piece rate treatments, the within-pair score gap is not significantly different across feedback rules.

This lower score of the underdogs can be due either to fewer submissions and/or to an increase in the number of mistakes. Intense feedback on relative performance can discourage the subject who is lagging behind, leading to fewer submissions. Or, it can generate a higher level of stress which leads to more mistakes. Which is the most important factor? Irrespective of the pay scheme the average percentage of mistakes of the underdogs is highest under the Continuous Feedback rule (18.92 % in PR-CF and 20.41 % in T-CF, compared with 12.45 % and 12.49 % in PR-NF and T-NF). In the Piece-Rate pay scheme the difference between NF and CF is significant ($p=0.015$), but not between NF and DF ($p=0.383$). Regarding submissions, the underdogs submit on average the same number of answers under each of the three feedback rules (84 in PR-NF, 83 in PR-DF, and 84 in PR-CF) ($p=0.760$ between NF and DF, $p=0.776$ between NF and CF). Hence the worse scores of the underdogs are mainly due to an increase in mistakes and not to a fall in submissions.

In the Tournament pay scheme the percentage of mistakes of the underdogs is also significantly higher ($p<0.001$) when feedback is continuous than when no feedback is given. Moreover, the average number of submissions by the underdogs is slightly lower when

¹⁰ Kolmogorov-Smirnov tests fail to find significant differences between the quantile distributions of score across treatments in the Piece-Rate pay scheme ($p>0.10$). But in the Tournament pay scheme the quantile distribution of T-NF differs from both T-DF ($p=0.095$) and T-CF ($p=0.10$). This is attributable to the lowest quartile subjects who perform less well in T-CF than in T-NF. This is consistent with the observation that the standard deviations of scores in T-DF (25.91) and T-CF treatments (23.04) are higher than in T-NF (17.12).

feedback is given (76.57 in T-DF and 76.96 in T-CF vs. 83.36 in T-NF), but these numbers are not significantly different ($p=0.338$ between NF and DF, $p=0.183$ between NF and CF). The underdogs' scores fall because mistakes crowd out correct answers, for (roughly) same number of submissions. These findings, which support Prediction 4, can be summarized as follows.

Result 4: Providing continuous feedback does not significantly reduce the underdogs' number of submissions, but it significantly increases the number of mistakes.

4.3 Peer Effects and the Dynamics of Submissions and Quality

We next analyze the dynamics of submissions and mistakes over time. Consider first the Piece Rate pay scheme. As stated in Prediction 2, positive (negative) peer effects would imply that a subject, who in the PR-DF treatment after 10 minutes learns that he is either trailing behind the other subject or is ahead of him, submits more (fewer) answers during the second half –10 to 20 minutes – relative to the submissions of similar subjects in the same period in PR-NF. Similarly, in PR-CF a subject who is repeatedly informed of his relative position should submit more (fewer) answers relative to a similar subject during the same period in PR-NF.

We construct a ratio relating the number of submissions in the last 10 minutes to the number of submissions during the first 10 minutes.¹⁶ There is no significant difference in this ratio between PR-NF and PR-DF ($p=0.458$) or PR-CF ($p=0.419$). This yields Result 5 which supports Prediction 1.

Result 5: There is no evidence of positive peer effects on the dynamics of submissions in the Piece Rate pay scheme.

Regarding tournaments, one of our hypotheses for the T-DF treatment is that, in the absence of peer effects, the average number of submissions in the second half will be decreasing in the

¹⁶ We acknowledge that this measure is somewhat arbitrary when used on the Continuous Feedback data, since under this rule subjects can continuously adjust their effort. However, it allows us to compare behavior under the Continuous and the Discrete Feedback rules.

score difference halfway through the production period (Prediction 3A). The larger this difference, the lower is both the front runner's and the underdog's number of submissions. It follows that an indication of positive peer effects in T-DF is a subject who, on having been informed on his relative position, either submits more answers (strong version of Prediction 3B) or does not submit significantly fewer answers (weak version of Prediction 3B) in the second half, compared to the submissions of similar subjects in the same period in T-NF. In T-CF, positive peer effects give rise to a similar pattern, where subjects do not submit significantly fewer answers than is observed in T-NF.¹⁷

Comparing the ratio of submissions in the last 10 minutes to submissions during the first 10 minutes, we find that there is no significant difference between T-NF (mean=1.068) and T-DF (mean=1.066; $p=0.559$), whereas the difference between T-NF and T-CF is close to significance (mean = 1.043; $p=0.105$).

These non-parametric statistics do not, however, control for the number of submissions during the first half of the period and for the score differentials. To account for this we estimate two linear regression models. In the first, the dependent variable is the ratio of the number of submissions in the last 10 minutes to their number in the first 10 minutes. In the second, the dependent variable is the corresponding ratio for mistakes. The explanatory variables include the score after 10 minutes, the subject's gender, and dummy variables for the Discrete and Continuous Feedback treatments, the No Feedback rule being the omitted reference category. We interact these variables with the competitor's gender and score after 10 minutes. We hypothesize that the difference in scores with the competitor influences the evolution of submissions, and allow this effect to differ for the genders. The models are estimated separately for the underdogs and the front runners; see Table 4.¹⁸

¹⁷ In T-CF, this effect could manifest itself earlier or later as the subjects have several opportunities to adjust their behavior. We analyze effort after 10 minutes to compare with T-DF. See previous footnote.

¹⁸ We estimated the same models for the Piece Rate treatments. The results are however omitted from Table 4

(Table 4 about here)

Consider first the underdogs. Their ratio of submissions is not affected by their own intermediate score: whatever their score in the first half of the experiment, they continue to submit answers with the same intensity in the second half of the game. The evolution of submissions does not differ in the Discrete Feedback and in the No Feedback treatments, and in the Discrete Feedback treatment the underdogs do not adjust their number of submissions to the score differentials after ten minutes. This is not consistent with the strong Prediction 3B but it is consistent with the weak version of Prediction 3B. Considering the T-CF treatment we find that, controlling for their score in the first half of the game, the underdogs adjust their submissions to the score gap, as predicted by Hypothesis 3A. The regression shows also however that, *ceteris paribus*, they increase their submissions significantly more in T-CF than in T-NF. This could be interpreted as some evidence of peer effects, provided the score gap is not too high. We acknowledge however that the statistical power of this regression is quite low.

For the Discrete and Continuous Feedback rules we find that, controlling for their score, front runners do not adjust their submissions to the distance with the underdogs. This is consistent with the weak Prediction 3B. This should, however, be qualified when feedback is continuous: when the underdog's score is higher, males' submissions increase less over time than that of females, but the magnitude of this effect is very small.¹⁹

Underdogs make significantly more mistakes in the second half of the game relative to the first in T-CF than in T-NF. Front runners with a higher score after 10 minutes decrease their rate of submissions and increase marginally their rate of mistakes in the second half of the game. An interpretation could be that these subjects have exerted an excessively high effort

since no coefficient was significant. This provides further support for Result 5 and Hypothesis 1.

¹⁹ This could be due to different attitudes toward risk (Eckel and Grossman, 2008) or to gender differences in overconfidence (Barber and Odean, 2001): female front runners may require more distance to the underdogs before they feel safe to reduce their own effort.

at the beginning of the experiment to maximize their probability to win the tournament and they are not able to work at the same pace in the second half of the game, possibly because of a fatigue effect.

These findings yield Result 6, which gives support to the weak Prediction 3B and to Prediction 4:

Result 6: In tournaments we find some evidence of a positive peer effects on the evolution of the underdogs' number of submissions and of a negative quality peer effect on the evolution of their number of mistakes.

4.4 Quitting

We define quitting as permanently stopping to work before the twenty minutes have elapsed. There is no rationale for such behavior in a piece rate pay scheme. But in a tournament, in the absence of peer effects, quitting is rational for an underdog when the likelihood of winning becomes sufficiently low. For front runners, it is also rational to quit if the underdog has dropped out.

13 subjects of the 14 who were behind halfway in the Tournament - Discrete Feedback treatment and 23 of the 24 involved in the Tournament - Continuous Feedback treatment eventually lost the competition. The losers of the tournaments are on average lagging behind their competitor during 92.26% of the twenty minutes. Yet, in all but one treatment subjects work until the 20th minute. In the T-CF treatment, only 3 subjects quit before the end of the game: one person quits at 14'08" and his co-worker responds by quitting at 16'59" (see pair 1 in Appendix 2), and 17'28" for another winner (see pair 2). Visually inspecting the dynamics of scores in each subject pair of subjects confirms that the subjects almost never drop out even when the outcome of the competition is highly predictable (see pair 3 for example). This also rejects Prediction 3A.

Result 7: In the Tournaments with Discrete and Continuous Feedback both underdogs and front runners almost never drop out, even when there seems to be little doubt about who will win.

Why do so few people quit? Being ahead may generate a positive peer effect raising feelings such as enthusiasm or joy (or spite) from being the best, and this may motivate the front runners to continue working, even at a high pace. Regarding the underdogs, as we do not elicit the subject's beliefs we cannot exclude that those who continued to work actually thought they might win. However, we do not find this a likely explanation: almost all underdogs continue to perform the task right to the end, even though the performance gap increases (see the performance graphs in Appendix 2). The lack of quitting is likely due partly to peer effects. The underdogs continue to work even if they are sure to lose, possibly in order to avoid unpleasant feelings, such as losing face by being known to give up. In our experiment there is no audience, but the underdog's competitor learns how much the competitor produced compared to his own score at the end of the game. There may also among underdogs be a desire to maintain a self-image as someone who never gives up and who always does one's best. Future work should seek to carefully disentangle these effects and evaluate their relative magnitude.

5. DISCUSSION AND CONCLUSION

We use a real-effort experiment to investigate the efficiency of various relative performance feedback systems under a piece rate and a tournament pay schemes. We find that releasing interim feedback on relative performance does not improve performance under either pay scheme. Under a piece-rate pay scheme agents' outcomes are not affected by performance comparisons. In tournaments with continuous feedback people adjust their submissions to the performance of others less than they would do if they were only motivated by monetary considerations. But since giving feedback is usually costly for the organization, our conclusion is, therefore, that it may be more profitable not to provide interim feedback on relative performance whatever the pay scheme in use.

On the other hand, if the agents are able to compare by themselves their own performance to that of their co-workers, the tournament pay scheme benefits from the presence of positive peer effects. Indeed, we have found that the underdogs do not give up even when the gap with their competitor is large enough to make the result of the tournament predictable, and consequently the front runners do not systematically slack off. Therefore, under these feedback rules peer effects make the tournament more efficient than expected.

The existence of peer effects has been documented for flat wage schemes (Falk and Ichino, 2006 and Mas and Moretti, 2009). We do not find evidence of positive peer effects when subjects are paid under a piece rate pay scheme. There are two possible explanations. First, it may be that positive peer effects are present but people in general work at their maximum level of effort under a piece-rate scheme, even under no feedback, and so positive peer effects due to feedback cannot manifest themselves in a strong and easily observable way. Our experiment provides some support for this being the case. Second, peer effects may be less prevalent in a piece rate pay scheme because this scheme focuses the attention on the direct monetary return from each additional unit of effort, leading people to disregard other dimensions of the environment such as performance comparisons. Our experiment cannot fully separate these two explanations, and further experimental work seem important.

Another important finding is based on the observation that while continuous feedback does not lead the front runners to slack off in tournaments, it significantly reduces the score of the underdogs. This negative impact, also observed in the piece-rate scheme, is not due to a reduction in their quantity of effort but to a decrease in the quality of their effort. We refer to this as a negative quality peer effect. This indicates that when an organization also cares about the performance and the quality of work of the less able agents, it should not implement a continuous feedback system, whatever the performance-pay mode. Our interpretation of this is that the output comparisons allowed by feedback generate stress and anxiety that

reduce the quality of effort. This is similar to how increased conscious attention to implicit competition, cash incentives, or the presence of an audience can reduce performance, as emphasized in the “choking under pressure” literature (Baumeister, 1984; Ariely et al., 2009; Dohmen, 2008). An alternative interpretation could be that mistakes are primarily due to the distraction or disruption created by intense feedback. That is, it is not the content of feedback but feedback itself that lowers quality. While continuous feedback can in general disrupt subjects' concentration we do not believe that it can explain the data, in particular the finding that feedback affects underdogs and front runners differently. These findings can on the other hand be explained by the quality peer effect.

Another noteworthy finding of our study is that dropping-out from the competition is rare. Both front runners and underdogs continue to exert high effort even when the distance between them is so large that it seems certain that the underdog cannot catch up, that is, when the marginal benefit of effort is null. We offer several explanations. First, people are motivated by the competition in itself, in which case the underdogs attempt to minimize the score difference relative to their competitor whereas the front runners try to maximize this difference. Second, as also discussed by Fershtman and Gneezy (2005), a social norm (‘one should never give up’, ‘always do your best’) may be at work. Third, ego (personal ambition activated by the joy of winning, self-esteem or self-image) is another source of motivation that could induce some subjects to attempt to maximize their score even when there are no more monetary incentives to do so.

Our study provides new evidence on the impact of feedback on relative performance on effort by using a particular task. Our laboratory experiment uses a specific and non-smooth substitution technology between performing the task and taking leisure on the job. Also, we captured the notion of the quality of effort in a very simple binary way. Of course, the output from many real life production processes are characterised by a much finer measure of

quality. It would therefore be useful to replicate our setup in other environments before drawing general conclusions. Many other extensions of our study can also be thought of, in particular manipulating the ability composition of groups to test more precisely whether more homogenous teams are better able to generate positive peer effects than heterogenous teams, or investigating the importance of peer effects in quitting decisions.

REFERENCES

- Aoyagi M. Information Feedback in a Dynamic Tournament. Osaka University Institute of Social and Economic Research Discussion Paper; 580; 2007.
- Ariely D., Gneezy U., Loewenstein G., Mazar N. Large Stakes and Big Mistakes. *Review of Economic Studies* 2009; 76(2); 451-469.
- Arvey R.D. Task Performance as a Function of Perceived Effort-Performance and Performance-Reward Contingencies. *Organizational Behavior and Human Performance* 1972; 8; 423-433.
- Bandiera O., Barankay I., Rasul I. Social Preferences and the Response to Incentives: Evidence from Personnel Data. *Quarterly Journal of Economics* 2005; 120 (3); 917-962.
- Barber B.M., Odean T. Boys Will Be Boys: Gender, Overconfidence, and Common Stock Investment. *Quarterly Journal of Economics* 2001; 116; 261-292.
- Baumeister R.F. Choking under Pressure: Self-consciousness and Paradoxical Effects of Incentives on Skillful Performance. *Journal of Personality and Social Psychology* 1984; 46(3); 610-620.
- Dohmen T. Do Professionals Choke under Pressure? *Journal of Economic Behavior and Organization* 2008; 65(3-4); 636-653.
- Eckel C., Grossman P. Forecasting Risk Attitudes: An Experimental Study Using Actual and Forecast Gamble Choices. *Journal of Economic Behavior and Organization* 2008; 68(1); 1-17.
- Ederer F. Information and Motivation in Dynamic Tournaments. MIT; Mimeo; 2008.
- Ederer F., Fehr E. Deception and Incentives: How Dishonesty Undermines Effort Provision? IZA Discussion Paper: Bonn; 3200; 2007.
- Ertac S. Social Comparisons and Optimal Information Revelation - Theory and Experiments. UCLA; Mimeo; 2005.
- Falk A., Ichino A. Clean Evidence on Peer Pressure. *Journal of Labor Economics* 2006; 24 (1); 39-57.
- Fershtman C., Gneezy U. Quitting. Tel-Aviv University Discussion Paper; 6; 2005.
- Gershkov A., Perry M. Tournaments with Midterms Reviews. *Games and Economic Behavior* 2009; 66 (1); 162-190.
- Gibbs M.J. An Economic Approach to Process in Pay and Performance Appraisals. Harvard Business School; Mimeo; 1991.
- Goltsman M., Mukherjee A. Interim Performance Feedback in Multistage Tournaments: The Optimality of Partial Disclosure. Mimeo; 2008.
- Greiner B. (2004). An Online Recruitment System for Economic Experiments. In: Kurt Kremer, Volker Macho (Eds), *Forschung und wissenschaftliches Rechnen 2003*. GWDG Bericht 63; Göttingen: Ges. für Wiss. Datenverarbeitung; 2004. p. 79-93.
- Gürtler O., Harbring C. Feedback in Tournaments under Commitment Problems: Theory and Experimental Evidence. IZA Discussion Paper: Bonn; 3111; 2007.
- Guryan J., Kroft K., Notowidigdo M. Peer effects in the workplace; Evidence from random groupings in professional golf tournaments. Forthcoming in *American Economic Journal: Applied Economics* 2009.

- Haraguchi K.H., Waddell, G.R. Interdependency in performance. IZA Discussion Paper: Bonn; 2944; 2007.
- Kandell E., Lazear E.P. Peer Pressure and Partnerships. *Journal of Political Economy* 1992; 100(4); 801-817.
- Kluger A.N., Denisi A. The Effects of Feedback Interventions on Performance: A Historical Review, a Meta-analysis and a Preliminary Feedback Intervention Theory. *Psychological Bulletin* 1996; 119; 254-284.
- Kräkel M. Emotions in Tournaments. *Journal of Economic Behavior and Organization* 2008; 67; 204-214.
- Lazear E.P. Pay Inequality and Industrial Politics. *Journal of Political Economy* 1989; 97; 561-580.
- _____. The Job as a Concept. In: W. Burns (Ed), *Performance Measurement, Evaluation, and Incentives*. Harvard Business School Press: Boston; 1992. p.183-215.
- _____, Gibbs M. *Personnel Economics in Practice*. John Wiley & Sons. Second edition; 2009.
- Lizzeri A., Meyer M., Persico N. The Incentive Effects of Interim Performance Evaluations. CARESS Working Paper; 02-09; 2002.
- Ludwig S., Lünser G. Knowing the Gap – Intermediate Information in Tournaments. University of Bonn; Mimeo; 2009.
- Mas A., Moretti E. Peers at Work. *American Economic Review* 2009; 99 (1); 112-145.
- Milkovich G.T., Newman J.M. *Compensation*. McGraw-Hill/Irwi: Chicago; 1996.
- Müller W., Schotter A. Workaholics and Drop Outs in Optimal Organizations. New York University; Mimeo; 2007.
- Niederle M., Vesterlund L. Do Women Shy away from Competition? Do Men Compete too Much? *Quarterly Journal of Economics* 2007; August; 1067-1101.
- Schotter A., Weigelt K. Asymmetric Tournaments, Equal Opportunity Laws, and Affirmative Action: Some Experimental Results. *Quarterly Journal of Economics* 1992; 107; 511-539.
- Yildirim H. Contests with Multiple Rounds. *Games and Economic Behavior* 2005; 51(1); 213-227.
- Zeiliger R. A Presentation of Regate, Internet Based Software or Experimental Economics. <http://www.gate.cnrs.fr/~zeiliger/regate/RegateIntro.ppt>., GATE; 2000.

Table 1. The treatments

Treatment	Payments	Feedback on co-player's score	# of sessions	# of subjects
Piece Rate – No Feedback (PR-NF)	€ 0.1 per unit output	After 20'	1	28
Tournament – No Feedback (T-NF)	€17 or €0	After 20'	2	44
Piece Rate – Discrete Feedback (PR-DF)	€ 0.1 per unit output	After 10' and 20'	1	30
Tournament – Discrete Feedback (T-DF)	€17 or €0	After 10' and 20'	1	28
Piece Rate – Continuous Feedback (PR-CF)	€ 0.1 per unit output	Throughout	1	30
Tournament– Continuous Feedback (T-CF)	€17 or €0	Throughout	2	48

Note: 'output' denotes the number of correctly added sets of numbers.

Table 2. Average and standard deviation of scores, mistakes, and submissions

Treatments	Score		Mistakes		Submissions	
	Mean	St. dev.	Mean	St. dev.	Mean	St. dev.
PR-NF (28)	84.00	20.79	10.86	7.63	94.86	19.52
PR-DF (30)	83.97	24.62	12.10	11.24	96.07	26.59
PR-CF (30)	81.50	24.40	13.30	7.25	94.80	20.88
Average (88)	83.14	23.15	12.11	8.87	95.25	22.35
T-NF (44)	80.30	17.12	9.59	6.28	89.89	15.02
T-DF (28)	75.93	25.91	11.36	5.84	87.29	26.03
T-CF (48)	76.33	23.04	12.17	6.61	88.50	21.67
Average (120)	77.69	21.72	11.03	6.37	88.72	20.54

Note: The number of observations by treatment is given in parentheses.

Table 3. Summary statistics on within-pair scores, mistakes, and submissions

Treatment	Average score		Wwithin-pair score difference	Average mistakes		Average submissions	
	Front runner	Underdog		Front runner	Underdog	Front runner	Underdog
PR-NF	94.14 (16.39)	73.86 (20.19)	20.28 (16.53)	11.21 (4.66)	10.50 (9.94)	105.35 (15.04)	84.36 (18.11)
PR-DF	95.20 (22.49)	72.73 (21.90)	22.47 (17.96)	13.93 (14.99)	10.27 (5.47)	109.13 (25.60)	83.00 (21.06)
PR-CF	95.00 (20.06)	68.00 (20.98)	27.00 (16.93)	10.73 (4.61)	15.87 (8.58)	105.73 (18.48)	83.87 (17.49)
Average	94.80 (19.42)	71.48 (20.72)	23.32 (17.00)	11.98 (9.42)	12.25 (8.40)	106.77 (19.90)	83.73 (18.53)
T-NF	87.64 (14.76)	72.95 (16.40)	14.68 (14.04)	8.77 (5.46)	10.41 (7.05)	96.41 (12.39)	83.36 (14.81)
T-DF	86.93 (23.78)	64.93 (23.84)	22.00 (19.01)	11.07 (5.88)	11.64 (6.01)	98.00 (24.69)	76.57 (23.45)
T-CF	91.42 (18.15)	61.25 (16.74)	30.17 (21.16)	8.63 (5.01)	15.71 (6.16)	100.05 (18.19)	76.96 (18.73)
Average	88.98 (18.29)	66.40 (18.91)	22.58 (19.26)	9.25 (5.39)	12.82 (6.81)	98.23 (17.90)	79.22 (18.61)

Note: Standard deviations in parentheses.

Table 4. Determinants of the evolution of submissions and mistakes in Tournaments

Dependent variable	Ratio of submissions in the last 10' to submissions in the first 10'		Ratio of mistakes in the last 10' to mistakes in the first 10'	
	Underdog	Front runner	Underdog	Front runner
Own score after 10 minutes	0.002 (0.002)	-0.006*** (0.001)	0.064*** (0.020)	0.028* (0.016)
Male	-0.007 (0.062)	0.041 (0.038)	0.454 (0.529)	0.330 (0.441)
Discrete Feedback treatment (DF)	-0.002 (0.162)	-0.031 (0.086)	1.265 (1.363)	1.837* (1.006)
Competitor's score after 10' in DF	0.000 (0.004)	0.001 (0.003)	-0.009 (0.031)	-0.044 (0.032)
Male*Competitor's score after 10' in DF	0.001 (0.002)	-0.001 (0.002)	0.000 (0.020)	-0.004 (0.021)
Continuous Feedback treatment (CF)	0.395*** (0.143)	0.002 (0.075)	2.592** (1.203)	0.514 (0.877)
Competitor's score after 10' in CF	-0.008*** (0.003)	0.000 (0.003)	-0.050* (0.026)	-0.031 (0.033)
Male*Competitor's score after 10' in CF	0.002 (0.002)	-0.005** (0.002)	0.010 (0.015)	-0.001 (0.021)
Constant	1.002*** (0.095)	1.322*** (0.061)	-1.116 (0.809)	0.066 (0.724)
Number of observations	60	60	58	59
Adjusted R ²	0.010	0.459	0.145	0.023



Notes: *, **, and *** indicate statistical significance at the 0.10, 0.05, and 0.01 level, respectively. Standard errors are in parentheses. We have fewer observations in the analysis of mistakes than in the analysis of submissions because 3 subjects did not make any mistake.

Appendix 1. The screen in the Piece Rate-Continuous Feedback treatment

(translated from French)

INCORRECT

Your score is the number of problems you have correctly solved. You have access to a total of 20 minutes.
Remaining time: **15 minutes, 43 seconds**

Your current score: 
Your co-participant's score: 

The numbers to add up are:

- 26
- 34
- 44
- 48

Enter your result:

[Submit](#)

Appendix 2. Evolution of performance over time in the Tournament-Continuous Feedback treatment

The figures below display the evolution of the scores of the two members of three different pairs of subjects minute by minute. The X-axis measures time (from the first to the 20th minute). The Y-axis measures the score. In the first figure below, the underdog stops working after 14 minutes, followed by the frontrunner after 16 minutes. In the second figure, the front runner stops working after 17 minutes but the underdog keeps working until the last minute. The third figure displays a much more frequent situation in which both competitors keep working until the very end of the game although the final outcome seems easy to predict at an early stage. The figures for all the other pairs are available from the authors upon request.

