

The Provision of Relative Performance Feedback: An Analysis of Performance and Satisfaction *

Ghazala Azmat

Queen Mary, University of London and Centre for Economic Performance (LSE)

Nagore Iriberry

*University of the Basque Country UPV/EHU, IKERBASQUE, Basque Foundation for
Research*

October 2014

Abstract

This paper studies the effect of providing relative performance feedback on individuals' performance, under two incentive schemes. In a laboratory setup, agents perform a real-effort task. We show that relative performance feedback increases performance when performance is related to pay (piece-rate) but has no effect on performance when pay is independent of performance (flat-rate). These effects are independent of the agent's relative position. Subjects are also asked to rate their satisfaction during the experiment. We find that under flat-rate, feedback has no effect on agents' satisfaction, while under piece-rate, feedback about relative position affects satisfaction.

Keywords: relative performance, feedback, piece-rate, flat-rate, satisfaction.

JEL classification: C91, M52, D04.

* We thank Manel Baucells, Michael Bashshur, Gary Charness, Ada Ferrer-i-Carbonell, George Loewenstein, Pedro Rey-Biel and anonymous referees for their comments. Ghazala Azmat acknowledges financial support from ECO2011-30323-C03-02, Fundación Ramon Areces and the support of the Barcelona GSE Research Network and the Government of Catalonia. Nagore Iriberry acknowledges financial support from Fundación Ramon Areces, Ministerio de Economía y Competición (ECO2012-31626), Departamento de Educación, Política Lingüística y Cultura del Gobierno Vasco (IT869-13) y Ministerio de Ciencia e Innovación (ECO2011-25295).

1. Introduction

Performance appraisals have become standard practice in organizations. Informing agents about how well they are performing relative to their peers—that is, providing workers with relative performance feedback—is a common method of implementing performance appraisals. Given its widespread use, it is important to understand the consequences of providing relative performance feedback. Managerial economics and social psychology have devoted quite a lot of attention to the study of performance appraisals (see Bretz et al., 1992, and Levy et al., 2004, for reviews). Research in economics, however, has paid little attention to relative performance feedback and its effects.

This paper studies the provision of relative performance feedback under two relevant incentive schemes: performance-related pay (piece-rate incentives) and fixed pay, independent of performance (flat-rate incentives). We examine whether there are differences in the response to feedback under the different payment mechanisms. In order to understand the extent of feedback's effectiveness, we study the influence of feedback on agents' emotional state (or level of satisfaction). Moreover, given that, in practice, feedback is rarely provided only once, we investigate the effects of its provision over time. By exploring all of these dimensions, we provide a comprehensive empirical analysis of the effects of relative performance feedback and study the channels through which it is effective.

We propose a controlled laboratory setup in which subjects perform a real-effort task and are rewarded according to piece-rate or flat-rate incentives. There are four working periods. Between periods, the control subjects are provided with their absolute performance, while the treated subjects are provided with their absolute performance *and* with the average performance in the session. Once feedback is provided, both control and treated subjects' satisfaction is elicited via questions about their happiness, arousal and dominance levels.¹

With respect to performance, under piece-rate incentives, we find that relative performance feedback had a strong and positive effect on individual performance, even after controlling for individual characteristics, such as ability. Those subjects who received relative performance feedback increased their performance by 17 percent compared to those who did not. In contrast, under flat-rate incentives, we find that the

¹ See Figure 1 for a graphical presentation of the experiment.

feedback does not affect subjects' performance. Interestingly, control subjects perform similarly under both incentive schemes, but relative performance feedback affects performance only when payment is related to performance. One may argue that when pay is related to performance, the information is "relevant," and individuals react because relative performance information has consequences in terms of relative income (i.e., strong effect under piece-rate but insignificant effect under flat-rate incentives).

With regard to the dynamic effects of providing relative performance feedback (under piece-rate incentives), we find that in all periods, the treated subjects outperform the untreated, although the effect becomes weaker over time. In addition, the *content* of the feedback (i.e., positive (negative) feedback when agents are informed that they are performing above (below) the average) does not affect subjects' subsequent performance differently since all subjects increase their performance.

When studying satisfaction, we find that the incentive scheme matters. Under piece-rate incentives, relative performance feedback had strong effects on both happiness and dominance levels. Contrary to the findings on performance, we show that the treatment had very different effects on those who received positive versus negative feedback. We find that receiving positive (negative) feedback affects subjects' happiness and dominance levels positively (negatively), such that when we consider only the overall treatment effect, the opposite signs cancel out. With respect to happiness, between those performing above and below the group average, the relative feedback leads to a seven-percentage-point increase in the gap (or inequality) of subjects' happiness. With respect to dominance levels, the increase in inequality is eight percentage points. Moreover, the inequality in both happiness and dominance increases over time with cumulative information. We do not see any effect of relative performance feedback on satisfaction under flat-rate incentives. This indicates that such feedback affects individuals' emotional state only when this information has consequences in terms of relative income (i.e., under piece-rate).

These findings on performance and emotional response are consistent with a competitive-preferences framework in which individuals care about income comparisons. Individuals get extra utility from knowing that they earn more than others and disutility when they learn that they earn less than others—i.e., individuals have competitive preferences.

The findings from this paper have important implications for understanding whether or not an organization would choose to provide relative performance feedback.

We show that there are very strong effects on performance when performance is linked to payment. This implies a strong incentive for an organization to employ this mechanism, especially because the cost to implement it is negligible. However, we go beyond the conventional thinking on this issue to highlight that this mechanism, while very effective at increasing performance, also has important consequences for individuals' affective state or satisfaction.

Understanding how agents feel when they receive relative performance feedback is important, as it sheds light on when the information is relevant to the agent. Bowles (1998) argues that “economic institutions structure the tasks people face and hence influence not only their capacities but their values and psychological functioning as well.” In recent years, along with more-traditional outcome variables, economists and managers have started to give more weight to individuals' happiness as an important outcome variable. Individuals' well-being has been linked with economic outcomes. For example, labor-market studies that include broader measures than happiness have found that workers' psyches are correlated with their morale and motivation, as well as with other relevant issues such as firms' retention ability (see Bewley, 1999).²

The paper is organized as follows. Section 2 includes a review of the existing literature. Section 3 describes the experimental design and procedures in detail. Section 4 provides a theoretical framework of thinking about the expected effects. In Section 5, we present the raw results via tables and figures. Sections 6 and 7 include the regression analysis, quantifying the main treatment effects under piece-rate and flat-rate incentives, respectively. Finally, Section 8 concludes.

2. Related Literature

The provision of feedback on individual effort—under either piece-rate or flat-rate incentives—has recently been studied in several environments. Under piece-rate, Hannan et al. (2008), using a laboratory experiment without real effort, Azmat and Iriberry (2010) in an educational setup, and Blanes i Vidal and Nossol (2011) in a firm setting, find that the provision of feedback has a positive effect on performance. In contrast, Eriksson et al. (2009), in the laboratory with real effort, find that although feedback does not affect performance, it increases the mistake rate of the worst-performing agent. Under flat-rate incentives, Kuhnen and Tymula (2012) study the

² Kahneman et al. (1999), Loewenstein (2000), Frey and Stutzer (2002), Krueger (2005) and McFadden (2005) provide recent overviews of research on happiness and its relation to economics.

effect of feedback in relation to individuals' expectations of their relative performance and find that subjects exert more effort and expect to rank better when they are told that they may learn their ranking.³ Our paper contributes to this literature on the effect of feedback on performance by explicitly comparing the effect of relative performance feedback under both piece-rate and flat-rate incentives, a comparison that the existing literature does not make. This allows us to disentangle individuals' responses to relative income versus relative performance comparisons.

In contrast with piece-rate or flat-rate incentives, under tournament incentives, relative performance feedback can be informative on the probability of winning the tournament, such that individuals are, indeed, expected to react to this information. Casas-Arce and Martinez-Jerez (2009), Young et al. (1993), Muller and Schotter (2010), Hannan et al. (2008), Fehr and Ederer (2007) and Eriksson et al. (2009) perform empirical tests on the effect of providing relative performance feedback in tournaments. The evidence is mixed. While some authors find that the provision of such feedback increases the effort of all participants, others find that the leading participants slack off, while those who are lagging behind give up. In our paper, the focus is only on piece-rate and flat-rate incentives, abstracting from a competitive-environment setting in which feedback is informative of one's own payoff.

It is important to note that in our setting, relative performance information is provided privately, which distinguishes it from settings in which pure peer effects or status-seeking effects play a role. Peer effects—that is, observing others' work and/or being observed—has shown to be effective in changing individuals' choice of efforts (see, for example, Falk and Ichino, 2006, and Mas and Moretti, 2009). When relative performance feedback is publicly available, status-seeking concerns may arise (see, for example, Charness et al., 2010, in which individuals can make investments to improve their own performance or sabotage others' performance).

The literature on the effects of feedback on affective response is relatively scarce.⁴ To understand how relative comparisons affect happiness, Charness and Grosskopf (2001) use a dictator-type experimental design to test whether there is a relationship between a person's happiness and the weight she attaches to relative payoffs. They do not find support for the hypothesis that happiness levels are inversely

³ Recently, Bandiera et al. (2009) and Delfgaauw et al. (2013) also considered the impact of relative performance feedback at the team level.

⁴ Kahneman et al. (1999), Loewenstein (2000), Frey and Stutzer (2002), Krueger (2005) and McFadden (2005) provide recent overviews of research on happiness and its relation to economics.

related to a preoccupation with relative payoffs. Brandts et al. (2009) study, in the laboratory, the effect of a competitive environment on happiness and find that competition leads to increased inequality in happiness between subjects who are chosen to participate and those who are excluded during the competition. In a recent field study, Card et al. (2012) find that disclosing information on peers' salaries affects their job satisfaction, such that workers with salaries below the median for their pay unit and occupation report lower job satisfaction, while those earning above the median report no higher satisfaction. Our paper is the first to look at the effect of relative performance feedback on agents' affective response in a controlled environment.

Overall, our study makes two important contributions to the literature. First, we go beyond the effect of relative performance feedback on individual performance by also analyzing the effect on individuals' affective response. Second, we compare the effect of relative performance feedback under two commonly used incentive schemes, piece-rate and flat-rate, allowing us to disentangle the effect of relative income feedback from that of relative performance feedback on both performance and affective response.

3. Experimental Design and Procedures

Sixteen experimental sessions were conducted in the Laboratori d'Economia Experimental (LEEX) at Universitat Pompeu Fabra, using z-Tree experimental software (Fischbacher, 2007). A total of 316 subjects, 20 per session, were recruited using the ORSEE recruiting system (Greiner, 2004), ensuring that subjects had not participated in similar experiments in our laboratory in the past. Between April and May 2009, we conducted eight sessions (160 subjects) under the piece-rate scheme, and in February 2010, the remaining eight sessions under the flat-rate scheme (156 subjects).⁵ A sheet with general, identical instructions was distributed and read aloud after all subjects had arrived. These instructions are in the Appendix. Subjects were guaranteed a three-euros show-up fee. Throughout the experiment, the subjects were seated in cubicles so that they could not observe other subjects' screens. Once the experiment had concluded, while they waited to be paid, subjects filled out a questionnaire that requested demographic details.⁶ We use these details as control variables in the econometric

⁵ All sessions had 20 participants, except for two sessions under flat-rate, where only 18 subjects showed up.

⁶ All sessions, treated and control, were mixed in terms of gender. Consistent with the gender composition of the subject pool at Universitat Pompeu Fabra, there were slightly more female than male subjects (39% men and 61% women).

analysis. The questionnaire can be found in the Appendix. Subjects were paid individually and in private, using a sealed envelope. Each experimental session lasted one hour.

Figure 1 shows the timeline of the experiment, which consisted of four working periods of five minutes each. During each period, subjects were presented with the same summations of four randomly generated numbers of two digits each, for which they were asked to submit an answer.⁷

Half of the subjects, 160, were rewarded according to their performance—that is, the number of correctly solved summations. More specifically, subjects were paid 0.15 euros for each correctly submitted answer (piece-rate incentives). All four working periods counted equally, and the total payment depended on the number of correctly solved summations during the four periods. (See Section 6 for the results under piece-rate.) The other 160 subjects were rewarded according to a fixed wage of 15 euros, which was independent of their performance (flat-rate incentives). (See Section 7 for the results under flat-rate).

At the end of each working period, four times during the experiment, subjects could rest for two minutes while they were given feedback. After the two minutes with the feedback information, they were asked to answer three questions included in the *Self-Assessment Manikin* (SAM) by Lang (1980). We will proceed to explain the feedback, as well as the SAM questionnaire.

The treatment variable of interest is the feedback about performance, which was provided at the end of each period. Half of the subjects (80 subjects under piece-rate and 78 subjects under flat-rate)—the control group—were provided with information about their correct number of summations. We refer to this as *absolute performance*. The other half (80 subjects under piece-rate and 78 subjects under flat-rate)—the treatment group—were provided not only with information about their correct number of summations, but also with the average correct number of summations within the experimental session. Therefore, subjects in the treatment group could observe whether they performed better or worse than the average, as well as their distance from this average. We refer to this information as the *relative performance* feedback. It is important to note that all subjects, both in the treatment and control groups, had

⁷ Many experimental papers have used summations as a real-effort task since they combine both ability and effort. See, for example, Niederle and Vesterlund (2007) and Eriksson et al. (2009).

received an explanation of the type of feedback they would get (see instructions in the Appendix).⁸

The SAM questionnaire, which is commonly used in social psychology, measures the affective response to a task. It consists of three sets of five pictures each, as shown in Figure 2. The top set measures happiness on a numerical scale between 1 and 9, where 1 represents *happy* and 9 represents *unhappy*. The second set measures arousal on a scale between 1 and 9, where 1 represents *aroused* and 9 represents *unaroused*. The third set measures dominance on a scale between 1 and 9, where 1 represents *dominated* and 9 represents *dominant*. In all administrations involving SAM, the subjects were instructed to rate their personal satisfaction using the pictures provided. The SAM instructions included the list of words from the pertinent end of each semantic differential scale in order to identify the anchors of each dimension to the subject. Thus, the subjects were instructed, for example, to use the extreme happy SAM rating if their reaction was one of feeling “happy, pleased, satisfied, contented, hopeful, relaxed,” and to use the other extreme if they felt “unhappy, annoyed, unsatisfied, melancholic, despairing, or bored.” Similar instructions accompanied all three scales.⁹

Given that the SAM questionnaire was administered immediately after providing subjects with feedback, we can be concerned that experimental demand effects drove any effect on individuals’ satisfaction. This means that subjects react to the feedback treatment because they think that the experimenter expects them to react. That is a pervasive concern without a clear solution in any study that elicits subjects’ satisfaction. We addressed this in two ways. First, instead of following a within-subject design, we followed a between-subject design, which is regarded as the safer methodology for avoiding demand effects. Second, we conducted our experiment under both piece-rate and flat-rate incentives and found radically different results depending on the incentives. These differences, which we discuss at length in Section 7, rule out the concern that the effects on satisfaction are due to pure demand effects.

⁸ Eight sessions were provided with relative performance feedback, and the other eight sessions were provided with absolute performance feedback. We had control and treated sessions run on the same weekdays and at the same times for each incentive scheme.

⁹ The subjects were instructed to use the extreme aroused SAM rating if they felt “stimulated, excited, frenzied, jittery, wide awake, aroused” and to use the other extreme if they felt “relaxed, calm, sluggish, dull, sleepy, unaroused.” The subjects were instructed to use the extreme dominated SAM rating if they felt “controlled, influenced, cared for, awed, submissive” and to use the other extreme if they felt “controlling, influential, in control, important, dominant.”

After the four working periods, the four periods of feedback and the SAM questions were completed, subjects were informed about their total earned money. The treatment group under piece-rate was also informed about the average earnings in the experiment, so that, again, they could observe whether they would get higher or lower pay than the average subject in the experimental session. The total average earnings in the experiment under piece-rate, including the show-up fee, were 14.10 euros, with the average being 13.45 euros in the control group and 14.74 euros in the treatment group. The earnings in the experiment under flat-rate, including the show-up fee, were 15 euros for both the control and treatment groups. We chose the 15-euro flat-rate as it was close to the average earnings under piece-rate.

4. Theoretical Framework

This section provides a framework for understanding the potential effects of feedback on both performance and emotional state under different incentive schemes. Our starting point, the null hypothesis, is that individuals should not respond to feedback since the underlying incentive schemes are held constant independent of this information. However, a number of alternative hypotheses can be considered.

First, if we consider a standard selfish preferences set-up, agents may react to relative performance feedback because they have an imperfect knowledge of their own ability, and the feedback is informative of one's ability. According to the self-perception ability hypothesis, where effort and ability are complements, agents who learn that they are performing above (below) the average would be encouraged (discouraged) when relative performance feedback is provided, compared to the situation in which this information is absent.¹⁰ Second, considering other-regarding-preferences, agents may have inherently competitive preferences, such that they get utility (disutility) from performing above (below) other agents. The provision of relative performance information allows agents to behave according to their competitiveness.¹¹ Third,

¹⁰ Ertac (2006) proposes a model in which ability and effort are complements in performance, predicting that individuals who learn that they are performing above (below) the average will exert more (less) effort than if they did not know. The psychology literature has also studied the role of self-assessment (see, for example, Sedikides, 1993).

¹¹ Many models that incorporate competitiveness have been proposed. Kandel and Lazear (1992) propose a model where *peer pressure* enters additively into the utility function. A specific form of peer pressure mentioned by the authors is the difference between the average effort and one's effort. Charness and Rabin (2002) propose a simple piece-wise linear utility in which others' payoffs affect one's utility. One type of interdependent preference their utility model includes is that of *competitive* preferences, where others' payoffs enter negatively in one's utility. Dubey and Geanakoplos (2004, 2005) and Moldovanu et al. (2005) assume that individuals have knowledge of the complete ranking, and they assume that

individuals may also interpret information about the average performance as the “norm” or the expected performance on a specific task, such that individuals want to comply with this norm and, thus, adjust their effort to meet the average performance.¹²

The three alternative hypotheses outlined above predict different reactions regarding performance, which we can test empirically. The first, the self-perception of ability hypothesis, predicts that high- (low-) performing agents will be encouraged (discouraged) and, thus, increase (decrease) their effort. The second, the competitive preferences hypothesis, predicts an overall increase in performance, as these preferences foster competition in which individuals want to improve on their relative position, irrespective of their initial position in the performance distribution. Finally, the norm hypothesis predicts a mean reversion result on performance, such that low- (high-) performing agents increase (decrease) their performance.¹³

Regarding the effect on emotional response, and considering that utility represents individuals’ emotions (Lowenstein, 2000), these alternative hypotheses rationalize an emotional reaction to the provision of relative performance feedback. According to the competitive preferences hypothesis or the self-perception ability hypothesis, agents who learn that they are performing above (below) the average will have a positive (negative) emotional reaction—that is, they reveal themselves to be happier, more aroused and more dominating (unhappier, less aroused and more dominated) compared to the situation in which they do not receive this information. Therefore, the provision of relative performance feedback will increase the inequality in the emotional response between high- and low-performing agents, compared to the situation in which this information is not provided. According to the norm hypothesis, agents will have a negative emotional response to the information unless they meet the norm, such that irrespective of performing above or below the average, their emotional response, compared with the no-information setup, will be negative.

Finally, we also compare responses under piece-rate and flat-rate incentive schemes. The different incentive schemes allow us to test for individuals’ underlying motivation—whether they respond to differences in performance and/or differences in income. Under piece-rate incentives, one cannot disentangle income from performance

individuals get positive utility from the number of individuals below them and negative utility from the number of individuals above them. Hopkins and Kornienko (2004) propose a utility in which “status” or position in the ranking enters multiplying the absolute income.

¹² See Elster (1989) for a discussion on how social norms affect economic behavior.

¹³ For a more detailed discussion of theoretical predictions, see Azmat and Iriberry (2010), Section 3.

information, while under flat-rate incentives, one can identify that the effect is driven only by performance information.

5. Overview of the Results

Table 1 presents the raw results for the effect of providing relative performance feedback. The main performance measure is *Correct*, which refers to the number of correct answers that subjects provide. We also look at the ratio of the number of correct answers to the number of submitted answers, which can be interpreted as the quality or effectiveness of performance, *Quality*. There are three aspects of satisfaction: subjective well-being (*Happiness*), feelings of arousal (*Arousal*), and feeling of dominance (*Dominance*). For all variables, we report the mean values for the control and treatment groups under both piece-rate and flat-rate incentives. We test for treatment differences using a Wilcoxon rank-sum test and report the p -values in columns 4 and 6.

The table shows that, under piece-rate, the feedback treatment has a positive effect on performance, where the effect is largest in the first period and then weakens over time. We do not find any effect of treatment on quality, with the exception of a positive effect in period 1. Looking by gender, we find that the (positive) performance differences are driven solely by male subjects. For all measures of affective response under piece-rate, we see that there is no significant effect of feedback. By gender, there are some notable differences in satisfaction: treated male subjects report being happier and more aroused than control male subjects, while treated female subjects report feeling less dominant than control female subjects. We also look at performance differences for other demographic differences (age, foreign, field of study and ability in math) between the control and treatment but, unlike gender, we do not find any significant differences.

Under flat-rate incentives, we do not find an effect on performance or its quality; nor is there any effect on measures of satisfaction. Therefore, comparing piece-rate to flat-rate incentives, we find that feedback treatment on performance is significant only when performance is rewarded (i.e., under piece-rate). Again, by gender, there are some interesting differences. Feedback treatment has a (small) positive effect on the performance of male subjects but a negative effect on the performance of female subjects. Treated female subjects are less aroused than control female subjects, and treated male subjects feel less dominant than control male subjects. As with piece-rate

performance, we do not see performance differences for other demographic characteristics.¹⁴

Figures 3 to 6 show the cumulative distribution functions for the treatment and control groups under piece-rate and under flat-rate incentives, confirming the previous results. Under piece-rate, the distributions in Figure 3 clearly show that the treated subjects outperform the control subjects. This difference is shown in all parts of the distribution, suggesting that the treatment had a positive effect on the treated subjects, independent of the information they were receiving—that is, whether they were performing above or below the average. Figure 4 further shows that the overall positive effect comes only from male subjects, while female subjects show no reaction. Under flat-rate, the distributions in Figure 4 show that the treated subjects do not behave differently from the control subjects. However, when separated by gender, as shown in Figure 6, men react positively to the information, while women react negatively.

In addition to the main results, it is important to mention the difference in the baseline performance under the two incentive schemes. We find that the control groups under piece-rate and flat-rate incentives perform similarly (p -value 0.4001). Although economic theory would suggest that agents would not exert effort when pay is independent of performance, empirical evidence from the lab suggests otherwise.¹⁵

A number of arguments for why individuals perform the work in the lab despite a flat-rate incentive scheme, have been put forward. First, as in our setting, there are no alternatives to doing the task, such as allowing subjects to surf the web or read magazines. Subjects may simply perform the task as a way to overcome “boredom.” Charness et al. (forthcoming), however, offer an alternative in a flat-rate setting but still find positive performance. Second, subjects may feel morally obliged to work, given that they will receive payment. This could be the underlying motivation for putting in the effort, despite having alternative tasks. Finally, the amount paid and the effort exerted under flat-rate might reflect the subjects’ reservation “wage.” In our setting, the

¹⁴ In the regression analysis in the next section, we test for differential treatment effects that depend on demographics, both under piece-rate and flat-rate incentive schemes. As in the raw analysis, other than gender, we do not find robust differential treatment effect. We only report the gender differentials in treatment. Other tables are available on request.

¹⁵ For example, Khunen and Tymula (2012) paid their subjects a flat-rate of 23 dollars and find that agents solved, on average, 10.58 multiplication problems in 90 seconds in the control treatment (no feedback). In another example, Charness et al. (forthcoming) paid a flat-rate of one euro per two minute period in which subjects could decode sets of one-digit numbers into letters from a grid of letters that is displayed on the computer screen. Subjects solved, on average, 23.54 decoding problems in the control group (no feedback).

flat-rate of 15 euros was set to match the average payment under piece-rate (14.10 euros). Offering a similar average payment under the two incentive schemes might explain the similarity in baseline performance.

6. Regression Analysis under Piece-Rate Incentives

This section consists of two main parts. First, we focus on the effect of relative performance feedback on individual performance. Second, we study the effect on satisfaction. For both performance and satisfaction, we start by measuring the overall effect of the feedback. We then look for gender differences and, finally, analyze the effect of the feedback content on each of the outcomes.

6.1. The Effect of Relative Performance Feedback on Performance

6.1.1. The Overall Effect

We estimate the following linear regression to quantify the average treatment effect on performance.

$$(1) Y_{it} = \alpha + \beta Treatment_i + \lambda Period_t + \delta X_i + \varepsilon_{it} ,$$

where the dependent variable, Y_{it} , refers to the performance measures, *Correct* and *Quality*, for individual i at period t . The variable *Treatment* identifies those who receive the relative performance feedback. We also include a time trend, *Period*. Finally, X_i is a vector of control variables, including gender, foreign, age and controls for ability. One concern may be that the treated group is of different ability than the control group. Although this is highly unlikely given the random allocation between treatment and control groups, we rule out this concern by using different measures of ability. First, we include a dummy for whether or not subjects' area of study was science-oriented (*Science*).¹⁶ Second, we include a dummy variable for whether the subject took the math test in the national-level university entry exam *Selectividad* (*Math_Test*).¹⁷

¹⁶ We classify a degree to be a *Science* degree if Mathematics plays an important role in that program. See the classification in the Notes of Table 2.

¹⁷ *Selectividad* is similar to the Scholastic Aptitude Tests (SAT) used in the United States, taken at the end of the fourth year of high school. *Selectividad* differs from the SAT in that it tests knowledge of the topics covered during the last year of high school. If the student wants to do a science-oriented degree, she must take the Math test in *Selectividad*. Otherwise, a student taking the Math test but pursuing an arts oriented degree is likely to be of high ability. The correlation coefficient between *Science* and *Math_Test* is 0.27. Eight subjects did not report this information, so we include a dummy variable *Math_Test_Missing*. The analysis without this dummy yields both qualitatively and quantitatively the same estimates. We also have information on subjects' grade in the math test. When we include the grade or grade distribution, the results do not change.

Columns 1 and 6 in Table 2 show the estimates for equation (1) for the correct summations and quality of performance, respectively. Since there are four working periods for each subject, we also weight the observations by the individual fixed effects (random-effects model). The treated subjects correctly answer 17 percent (treatment coefficient 2.06) more summations than the control subjects. Clearly, subjects increased their performance due to the feedback. However, the quality of performance does not change. This is due to the increase in the two variables—the number of submitted and correct answers—being very similar.

Time also matters, so we further investigate the treatment effect separately for each period. In columns 2-5 and 7-10 of Table 2, we estimate, by OLS, equation (1) for periods 1 to 4 for correct summations and quality of summations, respectively. Regarding the correct number of summations, over time, subjects improve in their performance, which implies that there is learning. Both control and treated subjects become accustomed to the computer application and the task.¹⁸ We can also see that this learning is steeper in the beginning and slows down in the last period for both control and treated subjects. More interestingly, treated subjects outperform the control subjects in each period. The effect is strongest in the initial period, suggesting an anticipation effect.¹⁹ The anticipation effect is also in line with those found by Blanes i Vidal and Nossol (2011) and Kuhnen and Tymula (2012). The quality of performance is significantly larger only in the first period, and, afterwards, the control and treated subjects do not differ in the quality of their performance. The coefficients in periods 2, 3 and 4 are not significantly different from one another.²⁰

In summary, the provision of relative performance feedback has a positive and significant effect on subjects' performance. Since the learning over time is the same for both treatment and control groups, and since we have also controlled for ability, neither of these components can explain the results. Therefore, feedback treatment explains the increase in performance.

¹⁸ Subjects were not allowed to use paper and pencil for summations. We observed that many subjects started doing the summations row by row instead of column by column. Many subjects quickly realized that doing the summations column by column and writing the answers by column was faster than otherwise, improving their performance substantially. We explore whether there is heterogeneity in learning for different demographic groups. We do not find any significant effect.

¹⁹ We observed the strong first-period effect in all four sessions for the treated subjects.

²⁰ When we combine all periods and compare the treatment effect for each period pair-wise, we find that they are not significantly different. For *Correct*, the *p*-values for periods 2 and 3, periods 2 and 4, and periods 3 and 4 are 0.949, 0.814 and 0.863, respectively. For *Quality*, the *p*-values for periods 2 and 3, periods 2 and 4, and periods 3 and 4 are 0.794, 0.791 and 0.996, respectively.

6.1.2. Gender Analysis

Many studies have shown gender to be an important variable when looking at behavior (see Croson and Gneezy, 2009). We investigate whether men and women react differently to the provision of feedback. We estimate the following equation:

$$(2) Y_{it} = \alpha + \beta Female_i + \chi Treatment_t + \delta Female_i * Treatment_t + \lambda Period_t + \gamma X_i + \varepsilon_{it} ,$$

where our variable of interest is the interaction between women and the treatment. The results are shown in Table 3: we find very strong gender effects. Note that we always control for demographics (*Foreign, Age, Science, Math_Test*) such that the strong gender differential is not being driven by differences between men and women in those characteristics. Treated women do not react differently from non-treated women, while treated men do react positively compared to non-treated men. Notice that from our estimates in column 1, on average, male subjects receiving feedback increase their performance by 5.62 relative to males in the control group, while treated females do not increase their performance. These results are striking since they imply that all of the observed effect on performance is attributed solely to men. While treated men increase their performance by 56 percent, treated women's performance does not change at all. We also find that there is a gender difference in the quality of the performance, such that treated male subjects make fewer mistakes compared with untreated male subjects, and there is no effect for female subjects. However, this gender difference in the quality of performance is significant only in period 1 and disappears.

6.1.3. Feedback Content

We now consider the importance of the content of the feedback. From periods 2 to 4, subjects can react not only to the provision of feedback, but also to the informational content in the previous periods. In other words, if a subject learns in one period that she is performing above (below) the average, that feedback might influence her performance in the subsequent period.

We distinguish between positive and negative feedback. From periods 2 to 4, positive (negative) feedback would imply that a subject performed above (below) the average in the previous period. We also look at the accumulation of feedback information. In periods 3 and 4, positive (negative) feedback would be a situation in which the subject either performed above (below) the average in the previous two

periods, or in which she has improved (worsened)—i.e., she was initially below (above) the average and is now performing above (below). Finally, in period 4, positive (negative) feedback would be a situation in which the subject has performed above (below) the average in the previous three periods, or in which she has improved (worsened).²¹ Note that for the treated subjects, this information is revealed—i.e., it becomes feedback—while for the control subjects it is not. For example, a control subject who has performed above the average will not be informed that she is performing above average. For simplicity, we refer to being above average as positive feedback, regardless of whether or not this information has been revealed.²²

We estimate the following regression:

$$(3) Y_{it} = \alpha + \beta Positive_{(t-1)i} + \chi Treatment_i + \delta Positive_{(t-1)i} * Treatment_i + \lambda Period_t + \gamma X_i + \varepsilon_{it} ,$$

where $Positive_{(t-1)}$ is an indicator that takes the value of 1 if the subject performed above the average, or improved over time (as explained above), and 0 otherwise. Y_{it} measures performance and quality of performance.

Panel A in Table 4 shows the estimates for equation (3) for the different periods. In column 1, consistent with our previous analysis, treated subjects outperform untreated subjects. In addition, subjects' performance is correlated from period to period, as shown in the coefficient *Positive*. Interestingly, the interaction between the treatment and *Positive* is not significant. This suggests that the feedback content is irrelevant for subsequent performances. In other words, what matters is the provision of feedback, and regardless of whether the feedback is positive or negative, performance will increase. Estimations for the different levels of cumulative feedback, in columns 2 and 3, imply similar results. The results are, however, somehow weaker given the limited number of observations as we approach the final period. With regard to the quality of performance, shown in columns 4 to 6, high-performing subjects have a higher quality of performance. In addition, consistent with previous findings, there is no

²¹ Improving includes the following two cases: (1) below the average in period 1, above the average in periods 2 and 3; and (2) below the average in periods 1 and 2, above the average in period 3. The reverse is true for worsening.

²² There are other ways of specifying positive and negative feedback. For example, one could use the *within* individual improvement (decline) over time, with respect to the average. In other words, although an individual may be consistently performing below the average (i.e., negative feedback in our specification), if, over time, she is getting closer to the average, then this may be perceived as positive feedback. We replicate our analysis using this specification, and the results are in line with those found using our specification.

significant treatment effect or differential treatment effect that depended on the content of the feedback.

Given the important gender differences we found in the previous section, it is interesting to understand whether men and women also react differently to the content of the feedback. Panel B shows the estimates for equation (3) for men only, while Panel C shows the estimates for women only.

From Panel B, we see that men react to the informational treatment, no matter whether they receive positive or negative feedback. This is consistent with the overall result. In Panel C, when we look at women, we find that the treatment, as well as the interaction of the content with treatment, is not significant. This is consistent with our previous finding that women do not react to either the informational treatment or the content of the feedback. Columns 2, 3, 5 and 6 give the estimates for the cumulative information. For men, the estimates suggest the same findings, but they become less significant due to fewer observations. Interestingly, for women, although not significant, there are some differential effects depending on the content of the information. The negative interaction term suggests that the response to positive feedback is smaller than the response to negative feedback.

We extend our analysis using a continuous measure of feedback content. Instead of positive (negative) feedback, we use the difference between one's performance, measured by *Correct*, and the average performance. We proceed in the same way and define the interaction between the treatment and the difference between one's performance and the average performance. We find qualitatively the same results—that is, the interaction between this difference and the treatment is insignificant.²³

Overall, in this section, we find three important results. First, the provision of relative performance feedback has a positive and significant effect on subjects' performance. In each period, the treated subjects outperform the untreated, although the effect gets weaker over time. Second, the actual content of the feedback (positive or negative) does not affect subjects' subsequent performance differently. This is consistent with the competitive preferences hypothesis, but not with the other two alternative hypotheses we considered, as they predicted different reactions depending on

²³ These results are available upon request.

whether the feedback was positive or negative. Third, we find a strong gender difference in the reaction to the treatment, with the overall effect driven solely by men.

6.2. Satisfaction: Happiness, Arousal and Dominance

6.2.1. The Overall Effect

Relative performance feedback potentially affects aspects other than performance. In particular, it may influence a subject's satisfaction. Organizations care about the affective state of their employees since it has been found that this affects relevant issues, such as workers' morale, firms' retention ability (Bewley, 1999), and even productivity (see, for example, Iaffaldano and Muchinsky, 1985; Warr, 1999; and Oswald et al., 2009). We measure three aspects of satisfaction: feelings of happiness (or subjective well-being), arousal and dominance.

We start by looking at the overall treatment effect on these measures. Equation (1) of section 6.1 is estimated, with the dependent variable Y_{it} now referring to happiness, arousal and dominance, respectively. The control variables, X_{it} , are the same as before. We also include an additional control that measures the number of correct summations, *Correct*.²⁴ It is reasonable to assume that performance and satisfaction measures are positively correlated (see Frey and Stutzer, 2002). Table 5 shows the results for the three variables of interest. The first column of each panel refers to the overall effect, controlling for time-fixed effects and weighting the observations by the individual fixed effects (random-effects model). The other four columns refer to periods 1 to 4, respectively.

The main result is that treatment is insignificant for all measures of satisfaction, implying that the provision of relative performance feedback does not affect the subjects' well-being, arousal and/or feeling of dominance. As expected, the number of correct summations is positively correlated with all three satisfaction measures. As for the other control variables, the only noteworthy finding is that, as one would expect, subjects find the task less interesting over time (arousal level decreases).

6.2.2. Gender

²⁴ We also estimate (1) without the variable *Correct*, and the main results remain the same.

In Section 6.1.2, we found that the provision of relative performance feedback affected men’s performance very differently from women’s. It is, therefore, natural to investigate whether there are gender differences in satisfaction.

We extend the analysis from the previous section, in line with equation (2), to look for gender differences in the treatment effect. Table 6 shows that the treatment is insignificant. Additionally, the interaction coefficient of female and the treatment implies that the provision of relative performance feedback does not affect men’s satisfaction differently from women’s.

6.2.3. Feedback Content

As with performance, we now study the effect of the informational content on the subjects’ satisfaction. We will estimate the following equation:

$$(4) Y_{it} = \alpha + \beta Positive_{it} + \gamma Treatment_i + \delta Positive_{it} * Treatment_i + \lambda Period_t + \chi X_i + \varepsilon_{it},$$

where Y_{it} refers to the satisfaction measures and $Positive_{it}$ is a dummy variable if the subject performed above the average or improved over time (as explained in section 6.1.3), and 0 otherwise. Note that unlike equation (3), the feedback content in the current period will affect the dependent variable in the same period (see the timeline of the experiment in Figure 1). This information is provided to the treated subjects just before eliciting their satisfaction. As with the performance analysis, we also consider how the cumulative feedback affects individual satisfaction.

Table 7 shows the estimation for the differential effects of the content of the feedback on satisfaction. In the first panel, columns 1 to 4, we consider the effect of the feedback content on happiness. As we saw in the previous estimates, people who are performing well—in this case, those who are performing above the average—reveal themselves to be significantly happier. More importantly, the treatment significantly increases the inequality between the happiness of subjects who are receiving positive feedback and those who are receiving negative feedback. We find that a treated subject who is performing below the average reports a lower level (-0.31) of happiness compared to an untreated subject performing below the average, whereas a treated subject who is performing above the average reports a higher level (0.49) of happiness compared to an untreated subject who is also performing above the average. This implies that control subjects performing above the average are seven-percent happier

than those performing below the average. Among treated subjects, those who are given positive feedback reveal themselves to be 15-percent happier than those who are given negative feedback. Overall, the informational feedback treatment leads to an increase in the inequality of subjects' happiness by eight percentage points.

This result is a key finding, and one that is not obvious when we only look at the average effect of the treatment on happiness, as the positive effect on those above average and the negative effect on those below the average are cancelled out. However, once we control for the content of the information, the treatment increases the difference in happiness between those who received positive and those who received negative feedback.

When we look at the cumulative effect of feedback, in columns 2 to 4, we see very similar results. The magnitude of the gap between the subjects performing above and below the average is increasing over time. In the final period, we find that in the control group, the gap in happiness between the above- and below-average performing subjects remains stable (from seven to 11 percent) for the treatment group, while the gap increases substantially with the cumulative feedback (from 15 to 21 percent).

Columns 5 to 8 show the results for arousal. Overall, we find no significant effect of the interaction between the content and the treatment. What we do see is that the subjects performing above the average, irrespective of the treatment, reveal themselves to be more aroused than those performing below the average. Also, over time, as expected, subjects' arousal goes down.

Columns 9 to 12 show the results for the feeling of dominance. Overall, we find a significant effect of the interaction between the content and the treatment. Providing subjects with positive feedback increases the gap in the feeling of dominance between those subjects performing above and those performing below the average. This implies that control subjects performing above the average feel four-percent more dominant than those performing below the average. For the treated subjects, however, those who receive positive feedback reveal themselves to feel 12-percent more dominant than those who receive negative feedback. Thus, the feedback treatment leads to an increase in the gap of subjects' dominance by eight percentage points. More importantly, when we look at the cumulative effects, we also find some interesting results. From columns 10-12, we see that consistently receiving positive (negative) feedback has a significantly positive (negative) effect on the treated subjects' feeling of dominance. In particular, in period 4, the feedback treatment leads to an increase in the gap of subjects' dominance

by 24 percentage points. We find that while, for the control group, the gap in dominance between the above- and below-average performing subjects remains stable (at four percent), for the treatment group, the gap increases substantially with the cumulative feedback (from 12 to 28 percent).

As in section 6.1.3, we also extended our analysis using a continuous measure of feedback content. Instead of positive (negative) feedback, we use the difference between one's performance and the average performance. We proceed in the same way and define the interaction between the treatment and the difference between one's performance and the average performance. We find qualitatively the same results—that is, the interaction between this difference and the treatment is positive and significant.²⁵

We also check for gender differences in the reaction to the feedback content and find no significant differences between men and women. This is in line with the results found in section 6.2.2.

This section shows that the treatment has very different effects on the satisfaction of those receiving positive versus negative feedback. We find that receiving positive (negative) feedback affects subjects' happiness and dominance levels positively (negatively), such that when we consider only the overall treatment effect, the opposite signs cancel out. This also suggests that the increase in happiness (dominance) for those subjects performing above average and the decrease in happiness (dominance) for those performing below average is of equal magnitude. Furthermore, the treatment increases the inequality in both the happiness and dominance levels. The increase in inequality is consistent with both self-perception of ability and the competitive preferences hypotheses. Given that we find a positive effect of performance feedback independent of its content, we conclude that the results are consistent with the competitive preferences framework.

Since we elicit subjects' satisfaction after the feedback treatment, we are concerned that there may be experimental demand effects. We are able to rule this out in the following section.

7. Analysis under Flat-Rate Incentives

We now study the effect of providing relative performance feedback under an incentive scheme that is independent of performance (i.e., flat-rate incentives). Under

²⁵ These results are available upon request.

piece-rate incentives, since agents are rewarded according to their absolute performance, the feedback informs them about both their relative *performance* and their relative *income*. It is important to understand whether individuals are reacting to either the relative performance or to the relative income, or to both. A good way to separate the two effects is by changing the underlying incentive scheme from piece-rate to flat-rate. Since, under piece-rate, what we observe is the net effect, the analysis under flat-rate allows us to disentangle the performance effect from the income effect.

We find no effect of providing relative performance feedback on either of the two measures of performance in the regression analysis in Table 8, confirming what we already observed in Table 1 and Figure 5. We have also analyzed pooling the two data sets and looked at the differential treatment effect under piece-rate and flat-rate incentives. We find that the interaction between the treatment and the incentive scheme is significantly different from zero.

There is, however, an interesting gender difference, as observed under piece-rate incentives. From Table 9, we can see that men react positively to the treatment, although the magnitudes are much smaller than under piece-rate incentives (19 percent versus 56 percent). Furthermore, we find that the interaction between treatment and females is negative (-13 percent) and significant, such that women react negatively to the relative performance information. Consistent with previous results, we do not find a strong effect on the quality of performance. The observed effects under piece-rate incentives measure the net effect of providing both relative income and relative performance feedback, while the observed effect under flat-rate incentives measures only the effect of providing relative performance feedback. From this comparison, we can conclude that, while both men and women react positively to relative income information, women and men react oppositely to the purely relative performance information.

Regarding satisfaction, we find that relative performance feedback has no effect on the three variables of interest. Contrary to what we found under piece-rate incentives, even when we control for the feedback content (learning that one is performing above or below the average), as shown in Table 10, we find that treatment is not significant.²⁶

The results suggests that the effects on both happiness and the feeling of dominance are a real response to *relevant* information and not driven by an

²⁶ As under piece-rate, we also check for gender differences in the reaction to the feedback content and do not find significant differences between men and women.

experimental demand effect. This is reassuring, as it strengthens the validity of our findings under piece-rate. We refer to it as relevant information because relative performance feedback under piece-rate has consequences in terms of relative income, while under flat-rate incentives it does not. These differences in the responses help us to also rule out that the effects on satisfaction found under piece-rate incentives are due to purely experimental demand effects.

We conclude that the provision of relative performance feedback is most effective under piece-rate incentives, as it pushes subjects to work harder and increase their productivity. Moreover, it has important effects for the subjects' satisfaction, increasing the inequality in happiness and the feeling of dominance between subjects who perform above and those who perform below the average, only under piece-rate incentives. In other words, when this information has consequences in terms of relative income, subjects react by being happier (unhappier) if they are above (below) the average because this translates into more (less) earnings than the average.

8. Conclusions

In this study, we have shown that relative performance feedback under piece-rate incentives is an important tool in increasing individuals' performance, independent of the feedback content. Given that the provision of this feedback is easy to implement and almost cost-free, it is an attractive policy to improve performance. However, individuals react only when relative performance information has consequences in terms of relative income (i.e., under piece-rate but not under flat-rate incentives). This implies that the incentive scheme under which the information is provided matters.

As further evidence of when the information matters, we investigate when individuals care about information via their levels of satisfaction. We show that relative performance feedback increases the inequality in individuals' happiness and feeling of dominance. Although enhancing performance is a positive result, the increased inequality in satisfaction might imply a possible trade-off to a principal or policy maker who is deciding whether or not to provide relative performance feedback. However, even if an organization places more weight on individual satisfaction than on performance per se, it is not entirely clear that it *would not* use relative performance feedback. For example, firms that care about the well-being of their workers may also choose to provide feedback since, by doing so, they may accelerate the match quality of their workers.

The findings in terms of performance and emotional response are consistent with a framework in which individuals behaviorally respond to the information on their relative position regarding income. In particular, individuals get extra-utility (disutility) if they are earning above (below) the average, such that this information fosters competition in which individuals want to improve on their relative position.

We also find a strong gender difference in the reaction to the treatment. In particular, we show that gender differences in performance are important when relative performance feedback is provided. Given that the provision of the feedback facilitates social comparison and that we observed that the increase in performance is consistent with the competitive preferences hypothesis, this finding can be related to the recent literature on gender differences in competitive environments (see Croson and Gneezy, 2009; Niederle and Vesterlund, 2011).²⁷ We show that the provision of relative performance feedback—which can promote competition among individuals but where competition is rather symbolic—is enough to create gender differences in performance.

The results of this paper open new research questions. For example, long-term effects of relative performance feedback need to be studied further—in particular, using field studies. Another important area of study is other behavioral responses to feedback, such as the option of quitting or taking actions that go against organizations' interests. Finally, further investigation into what drives the gender difference would be an important next step.

²⁷ Women are found to shy away from competition, showing a preference for non-competitive environments (Deaner, 2006a and 2006b; Niederle and Vesterlund, 2007; and Niederle and Yestrumskas, 2008), although this is reversed when we switch to matrilineal societies (Gneezy et al., 2007). Also, women tend to underperform in competitive environments compared to men, mostly because men's performance increases when competing against women (Gneezy et al., 2003; Gneezy and Rustichini, 2004; Antonovics et al., 2009; and Hogarth et al., 2009).

References

Antonovics, K., Arcidiacono, P. and R. Walsh. 2009. "The Effects of Gender Interactions in the Lab and in the Field." *Review of Economics and Statistics*, February 2009, 91(1), pp. 152-163.

Azmat, G. and N. Iriberry. 2010. "The Importance of Relative Performance Feedback Information: Evidence from a Natural Experiment using High School Students." *Journal of Public Economics*, volume 94 (7-8), pp. 435-452.

Bandiera, O., Barankay, I. and I. Rasul. 2009. "Team incentives: Evidence from a Field Experiment." Mimeo.

Bewley, T. F. 1999. "Why wages don't fall during a recession." Harvard University Press, Cambridge.

Blanes i Vidal, J. and M. Nossol. 2011. "Tournaments without Prizes: Evidence from Personnel Records." *Management Science* 57:1721-1736.

Bowles, S. 1998. "Endogenous Preferences: The Cultural Consequences of Markets and other Economic Institutions." *Journal of Economic Literature* 36, 75-111.

Brandts, J., Riedl, A. and F. van Winden. 2009. "Competitive Rivalry, Social Disposition, and Subjective Well-Being: An Experiment," *Journal of Public Economics* 93, 11/12, 1158-1167.

Bretz, R. D., Milkovich, G. T. and W. Read. 1992. "The Current State of Performance Appraisal Research and Practice: Concerns, Directions, and Implications." *Journal of Management* 18, 312-352.

Burchett, R., Willoughby, J., 2004. "Work productivity when knowledge of different reward systems varies: report from an economic experiment." *Journal of Economic Psychology* 25, 591-600.

Card, D., Mas, A., Moretti, E. and E. Saez. 2012. "Inequality at Work: The Effect of Peer Salaries on Job Satisfaction," *American Economic Review*, October, Vol. 102, No. 6.

Casas-Arce, P., and F. A. Martinez-Jerez. 2009. "Relative Performance Compensation, Contests, and Dynamic Incentives." *Management Science* 55, pp. 1306-1320.

Charness, G. and B. Grosskopf. 2001. "Happiness and Relative Payoffs - An Experimental Study." *Journal of Economic Behavior and Organization*, 45-3, pp. 301-328.

Charness, G., Masclet, D., and M.C. Villeval. Forthcoming. "Competitive Preferences and Status as an Incentive: Experimental Evidence." *Management Science*.

Clark, A., Frijters, P. and M. Shields. 2008. "Relative Income, Happiness and Utility: An Explanation for the Easterlin Paradox and Other Puzzles." *Journal of Economic Literature* 46-1, pp.95-144.

- Clark, A. E., Masclet, D., and M.C. Villeval. 2010. "Effort and Comparison Income: Experimental and Survey Evidence." *Industrial and Labor Relations Review*, Vol.63, no.3, pp.407-426.
- Croson, R. and U. Gneezy. 2009. "Gender Differences in Preferences." *Journal of Economic Literature* 47(2), 1-27.
- Deaner, R. O. 2006a. "More males run relatively fast in U.S. road races: Further evidence of a sex difference in competitiveness." *Evolutionary Psychology* 4: 303-314.
- Deaner, R. O. 2006b "More males run fast: a stable sex difference in competitiveness in U.S. distance runners." *Evolution and Human Behavior* 27: 63-84.
- Delfgaauw, J., Dur, R., Sol J. and W. Verbeke. 2013. "Tournament Incentives in the field: Gender Differences in the Workplace." *Journal of Labor Economics* 31(2).
- Duesenberry, J.S. 1949. "Income, Saving and the Theory of Consumer Behaviour." Harvard University Press, Cambridge.
- Easterlin, R. A. 1974. "Does economic growth improve the human lot? Some empirical evidence." David PA, Reder MW (eds.) Nations and households in economic growth.
- Elster, J. 1989. "Social Norms and Economic Theory," *Journal of Economic Perspectives*, 3 (4): 99–117.
- Eriksson, T., Poulsen, A. and M. Villeval (2009) "Feedback and Incentives: Experimental Evidence", *Labour Economics*, 16, pp. 679-688.
- Falk, A., and A. Ichino. 2006. "Clean Evidence on Peer Effects." *Journal of Labor Economics* 24(1): 39–57.
- Fehr, E. and F. Ederer. 2007. "Deception and Incentives: How Dishonesty Undermines Effort Provision." IZA Discussion Paper 3200.
- Festinger, L. 1954. "A theory of social comparison processes." *Human Relations* 7: 117-140.
- Fischbacher, U. 2007. "z-Tree: Zurich Toolbox for Ready-made Economic Experiments." *Experimental Economics* 10(2), 171-178.
- Frey, B. S. and A. Stutzer. 2002a. "What Economists Can Learn from Happiness Research." *Journal of Economic Literature*, vol. XL, 402-435.
- Frey, B. S. and A. Stutzer. 2002b. "*Happiness and Economics*." Princeton and Oxford: Princeton University Press ISBN O-69 1-06998-O.
- Gächter, S., Thöni, C. 2010. "Social comparison and performance: experimental evidence on the fair wage-effort hypothesis." *Journal of Economic Behavior and Organization* 76 (3), 531–543.

- Gneezy, U., Niederle M., and A. Rustichini. 2003. "Performance in Competitive Environments: Gender differences." *Quarterly Journal of Economics*, pp. 1049-1074.
- Gneezy, U. and A. Rustichini. 2004. "Gender and Competition at a Young Age." *American Economic Review Papers and Proceedings*, pp 377-381.
- Gneezy, U, Leonard, K. L. and J. A. List. 2008. "Gender Differences in Competition: Evidence from a Matrilineal and a Patriarchal Society." *Econometrica* 77-3, pp 909-931.
- Greiner, B. 2004. "The Online Recruitment System ORSEE 2.0 - A Guide for the Organization of Experiments in Economics." University of Cologne WP Series in Economics 10.
- Greiner, B., Ockenfels A., Werner, P. 2011. "Wage transparency and performance: A real-effort experiment." *Economics Letters* 111, pp 236–238.
- Hannan, R. L., R. Krishnan and D. Newman. 2008. "The Effects of disseminating Relative Performance Feedback in Tournament Versus Individual Performance Compensation Plans." *The Accounting Review*, 83-4.
- Hogarth, R. M., Karelaia, N. and C. A. Trujillo. 2009. "Under-achievement and the glass ceiling: Evidence from a TV game show." Mimeo.
- Iaffaldano, M. T. and P. M. Muchinsky. 1985. "Job Satisfaction and Job Performance: A Meta-Analysis." *Psychological Bulletin* 97-2, pp. 251-73.
- Kahneman, D., Diener, E. and N. Schwartz (Eds.) .1999. "Well-Being. The Foundations of Hedonic Psychology." Russell Sage Foundation, New York.
- Krueger, A. B. 2005. "Well-Being and Policy Evaluation.", Presentation at the Econometric Society World Congress 2005, London, August 20, 2005.
- Kuhnen, C. M. and A. Tymula. 2012. "Rank Expectations, Feedback and Social Hierarchies," *Management Science* 58 (1): 94-113.
- Lang, P. J. 1980. "Behavioral treatment and bio-behavioral assessment: computer applications." In J. B. Sidowski, J. H. Johnson, & T. A. Williams (Eds.), *Technology in mental health care delivery systems* (pp. 119-137). Norwood, NJ: Ablex.
- Levy, P. E. and J. R. Williams. 2004. "The Social Context of Performance Appraisal: A Review and Framework for the Future." *Journal of Management* 30, 881-905.
- Loewenstein, G. 2000. "Emotions in Economic Theory and Economic Behavior", *American Economic Review, Papers and Proceedings*, 90, pp 426-432.
- Mas, A. and E. Moretti. 2009. "Peers at Work." *American Economic Review*, 99(1).

- Müller, W. and A. Schotter. 2010. "Workaholics and Drop Outs in Optimal Organizations." *Journal of the European Economic Association* 8(4):717–743.
- Murphy, K. R., and J. N. Cleveland. 1991. "Performance appraisal: An organizational perspective." Boston: Allyn and Bacon.
- McFadden, D. 2005. "The New Science of Pleasure." Frisch Lecture, Econometric Society World Congress 2005, London, August 20, 2005.
- Niederle, M. and L. Vesterlund (2007) "Do Women Shy away from Competition? Do Men Compete too Much?." *Quarterly Journal of Economics*, 122-3, pp 1067-1101.
- Niederle, M. and A. H. Yestrumskas. 2008. "Gender Differences in Seeking Challenges: The Role of Institutions", Working Paper 13922.
- Oswald A., Proto, E. and D. Sgroi. 2009. "Happiness and Productivity." IZA Discussion Paper No. 4645.
- Sedikides, C. 1993. "Assessment, Enhancement, and Verification Determinants of the Self-evaluation Process." *Journal of Personality and Social Psychology* 65(2), pp. 317-338.
- Warr, P. 1999. "Well-being and the Workplace" in Kahneman, Diener and Schwarz, pp. 392-412.
- Young, S. M., Fisher, J. and T. M. Lindquist. 1993. "The effect of intergroup competition and intragroup cooperation on slack and output in a manufacturing setting." *The Accounting Review* 68 (3): 466-483.

Figures and Tables

Figure 1. Timeline of the Experiment

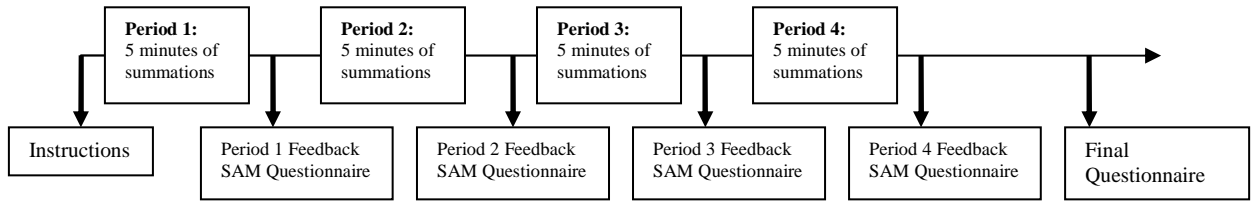
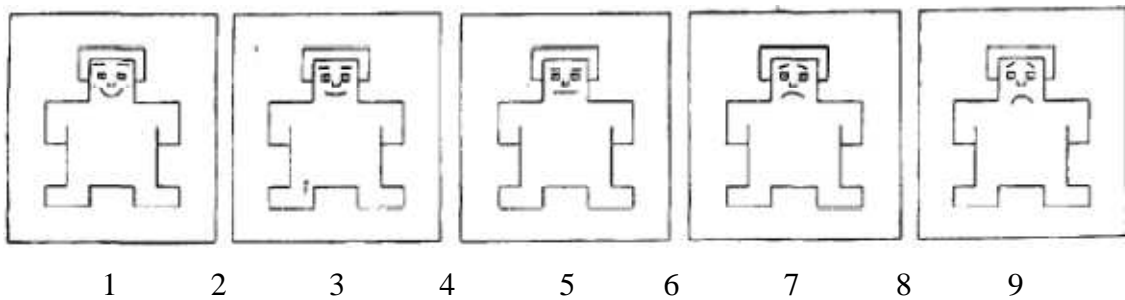


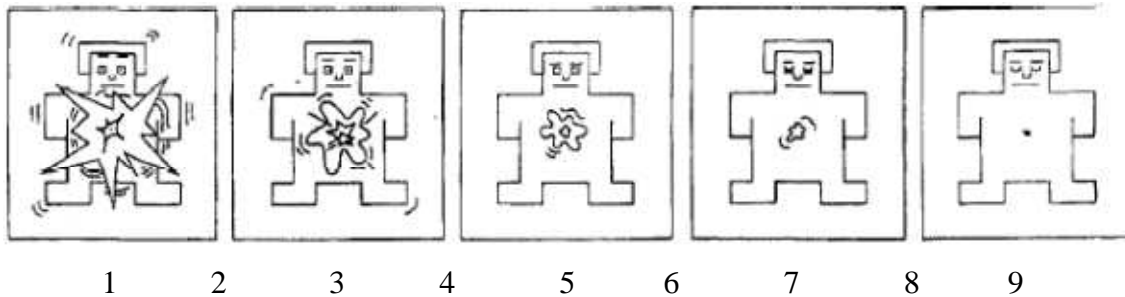
Figure 2. Self-Assessment Manikin (SAM)

Panel A: Happiness



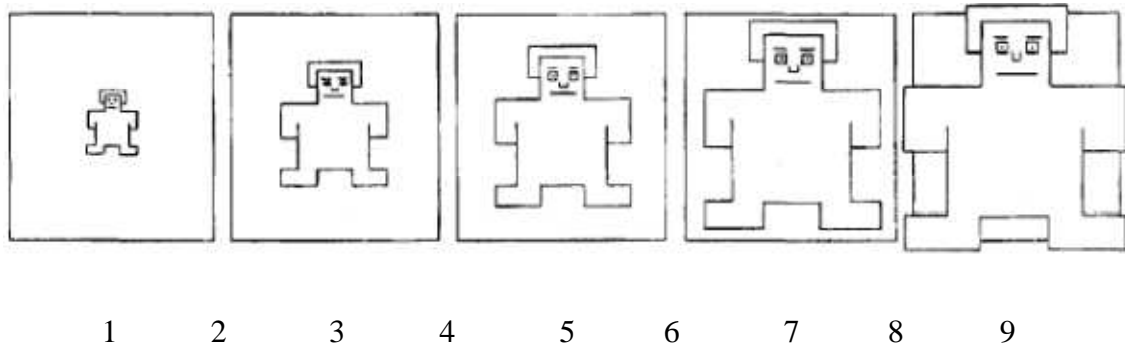
Note: 1 represents feeling “happy, pleased, satisfied, contented, hopeful, relaxed,” and 9 represents feeling “unhappy, annoyed, unsatisfied, melancholic, despairing, or bored.”

Panel B: Arousal



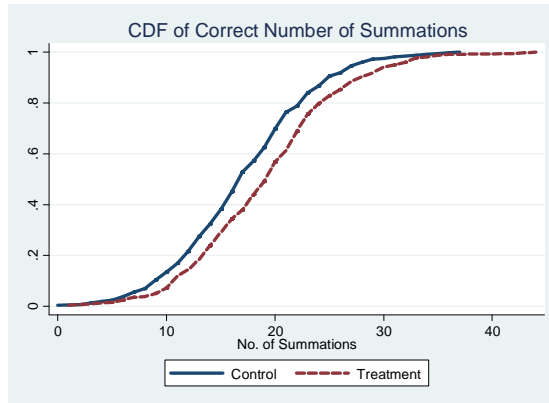
Note: 1 represents feeling “stimulated, excited, frenzied, jittery, wide awake, aroused,” and 9 represents feeling “relaxed, calm, sluggish, dull, sleepy, unaroused.”

Panel C: Dominance



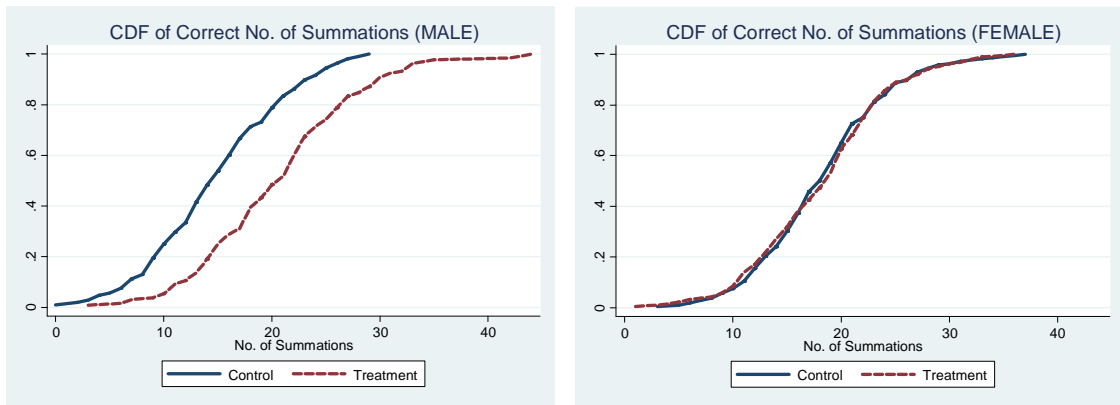
Note: 1 represents feeling “controlled, influenced, cared for, awed, submissive,” and 9 represents feeling “controlling, influential, in control, important, dominant.”

Figure 3: Cumulative Distribution under Piece-Rate



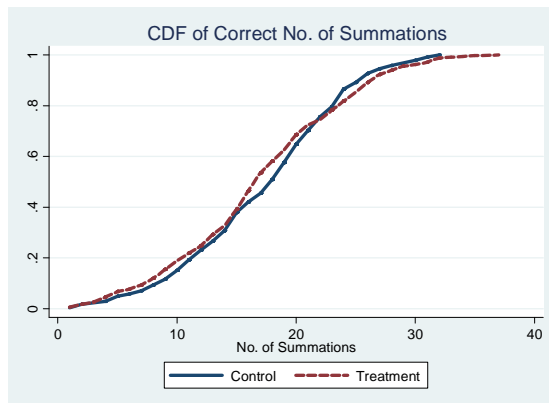
Notes: The p -value for the Kolmogorov–Smirnov test, comparing the distributions of the treated and control subjects, is given by 0.001.

Figure 4: Cumulative Distributions under Piece-Rate by Gender



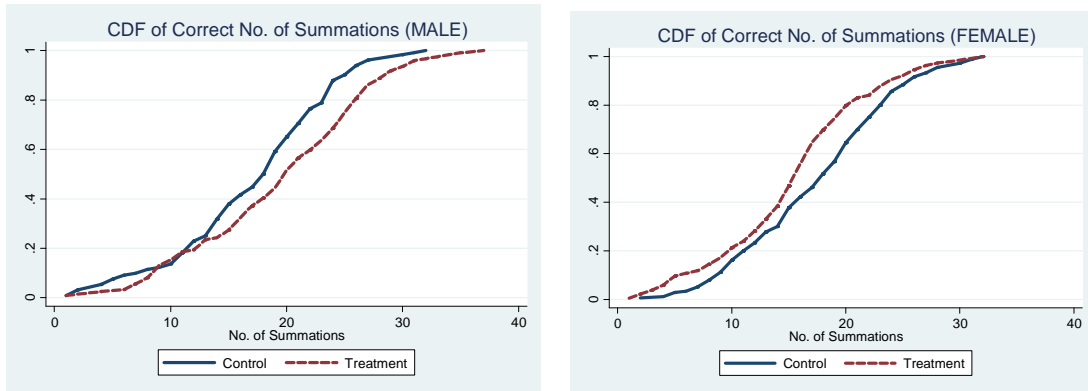
Notes: The p -value for the Kolmogorov–Smirnov test, comparing the distribution of the treated and control subjects, is given by 0.000 and 0.981 for male and female subjects, respectively.

Figure 5: Cumulative Distribution under Flat-Rate



Notes: The p -value for the Kolmogorov–Smirnov test, comparing the distributions of the treated and control subjects, is given by 0.238.

Figure 6: Cumulative Distributions under Flat-Rate by Gender



Notes: The p -value for the Kolmogorov–Smirnov test, comparing the distribution of the treated and control subjects, is given by 0.012 and 0.003 for male and female subjects, respectively.

Table 1: Raw Results by Treatment						
	Piece-Rate Incentives			Flat-Rate Incentives		
	Control	Treatment	<i>p</i> -value	Control	Treatment	<i>p</i> -value
Correct						
Overall	17.43	19.57	0.00	17.65	17.32	0.42
Period 1	13.39	16.51	0.00	14.15	13.60	0.41
Period 2	16.86	18.81	0.06	17.01	17.47	0.67
Period 3	19.38	21.24	0.08	19.37	18.54	0.38
Period 4	20.10	21.73	0.07	20.06	19.65	0.61
Male	15.12	21.03	0.00	17.44	19.77	0.01
Female	18.61	18.55	0.92	17.81	15.70	0.00
Quality						
Overall	0.85	0.87	0.14	0.87	0.85	0.22
Period 1	0.80	0.86	0.07	0.84	0.84	0.72
Period 2	0.86	0.87	0.54	0.87	0.85	0.98
Period 3	0.87	0.87	0.66	0.89	0.84	0.06
Period 4	0.87	0.87	0.93	0.87	0.86	0.84
Male	0.81	0.88	0.00	0.85	0.86	0.69
Female	0.87	0.86	0.40	0.88	0.84	0.04
Happiness						
Overall	6.66	6.60	0.86	6.71	6.65	0.87
Period 1	6.36	6.44	0.50	6.72	6.82	0.30
Period 2	6.63	6.48	0.76	6.55	6.71	0.60
Period 3	6.74	6.53	0.69	6.83	6.45	0.14
Period 4	6.93	6.95	0.68	6.76	6.64	0.82
Male	6.32	6.65	0.04	6.38	6.67	0.42
Female	6.83	6.56	0.17	6.96	6.64	0.42
Arousal						
Overall	4.86	5.07	0.10	4.81	4.55	0.08
Period 1	4.86	5.10	0.35	4.73	4.56	0.56
Period 2	5.14	5.20	0.92	4.91	4.63	0.33
Period 3	5.01	5.19	0.46	4.87	4.59	0.36
Period 4	4.41	4.80	0.18	4.75	4.40	0.29
Male	4.46	5.11	0.07	4.95	4.97	0.84
Female	5.00	5.05	0.43	4.72	4.27	0.01
Dominance						
Overall	6.32	6.28	0.34	6.41	6.10	0.02
Period 1	6.10	6.20	0.71	6.50	6.14	0.22
Period 2	6.36	6.21	0.25	6.32	6.21	0.63
Period 3	6.41	6.28	0.56	6.42	5.86	0.04
Period 4	6.40	6.44	0.91	6.39	6.19	0.38
Male	6.22	6.55	0.33	6.89	6.23	0.00
Female	6.37	6.10	0.03	6.05	6.02	0.90

Notes: The columns *Control* and *Treatment* provide the average performance (*Correct*), quality of performance (*Quality*), and emotional response (*Happiness*, *Arousal*, *Dominance*) for each group, respectively. The *p*-values refer to the Wilcoxon rank-sum test, which tests for the difference between the groups.

Table 2: Treatment Effect on Performance under Piece-Rate: Correct and Quality

	Correct					Quality				
	ALL	PERIOD 1	PERIOD 2	PERIOD 3	PERIOD 4	ALL	PERIOD 1	PERIOD 2	PERIOD 3	PERIOD 4
Constant	12.0878** [4.823]	11.9685** [5.678]	16.7831*** [5.138]	22.6080*** [4.934]	17.3480*** [5.351]	0.8426*** [0.084]	0.8797*** [0.205]	0.8728*** [0.113]	0.9203*** [0.100]	0.8425*** [0.093]
Treatment	2.0601** [0.902]	3.0739*** [1.005]	1.8354* [0.984]	1.7255* [0.984]	1.6058 [1.008]	0.0198 [0.016]	0.0627** [0.026]	0.009 [0.023]	0.0016 [0.019]	0.0059 [0.016]
Period	2.0356*** [0.122]					0.0145*** [0.004]				
Female	0.4984 [0.982]	0.7143 [1.052]	1.0646 [1.059]	0.3408 [1.109]	-0.1262 [1.075]	0.0208 [0.016]	0.0516* [0.030]	0.0149 [0.022]	0.0235 [0.021]	-0.0067 [0.018]
Foreign	0.561 [2.024]	0.4755 [2.358]	1.6187 [2.049]	0.4245 [2.024]	-0.2745 [2.538]	0.0336 [0.031]	0.0478 [0.072]	0.0641* [0.038]	0.016 [0.044]	0.0066 [0.035]
Age	-0.0737 [0.220]	-0.0365 [0.265]	-0.1228 [0.231]	-0.2229 [0.215]	0.0876 [0.242]	-0.0027 [0.004]	-0.0072 [0.010]	-0.0021 [0.005]	-0.0033 [0.005]	0.0016 [0.004]
Science	0.2636 [0.970]	0.104 [1.029]	0.4042 [1.001]	0.4266 [1.100]	0.1197 [1.114]	0.002 [0.016]	0.0042 [0.024]	-0.0003 [0.021]	0.0252 [0.022]	-0.0209 [0.018]
Math_Test	1.8610* [1.054]	2.2950* [1.252]	2.3779** [1.158]	1.4743 [1.181]	1.2969 [1.164]	0.0115 [0.018]	0.0307 [0.035]	0.0218 [0.028]	-0.0094 [0.022]	0.0031 [0.020]
Math_Test_Missing	-3.1609 [2.116]	-4.0274 [2.488]	-4.2998* [2.318]	-2.6649 [2.424]	-1.6517 [2.461]	0.0251 [0.030]	0.0361 [0.067]	-0.0121 [0.045]	0.0457 [0.041]	0.0306 [0.037]
Observations	640	160	160	160	160	640	160	160	160	160
Number of subject	160					160				
R-squared		0.088	0.061	0.049	0.025		0.084	0.022	0.026	0.022

Notes: Robust standard errors at the session level shown in brackets. * denotes significance at the 10% level, ** denotes significance at the 5% level and *** denotes significance at the 1% level. The variable *Science* takes value 1 if the degree is in Architecture, Biology, Business, Computer Science, Economics, Engineering or Human Resources and 0 if the degree is in Communication, Health, Human Science, Law, Marketing, Photography, Political Science and Translation and Interpretation. The variable *Math_Test* takes value 1 if the subject took the Math test in the national-level university entry exam “Selectividad.” *Math_Test_Missing* takes value 1 if the subject does not report whether he or she took the Math test in Selectividad (note that there are only eight subjects for whom this information is missing).

Table 3: Gender Differences in the Treatment Effect on Performance under Piece-Rate

	Correct					Quality				
	ALL	PERIOD 1	PERIOD 2	PERIOD 3	PERIOD 4	ALL	PERIOD 1	PERIOD 2	PERIOD 3	PERIOD 4
Constant	9.9488** [4.450]	9.3437* [5.190]	14.5116*** [4.912]	20.8086*** [4.632]	15.4876*** [5.122]	0.8103*** [0.079]	0.8104*** [0.192]	0.8534*** [0.121]	0.9025*** [0.098]	0.8195*** [0.088]
Treatment	5.6179*** [1.511]	7.4395*** [1.609]	5.6135*** [1.720]	4.7184*** [1.747]	4.7001*** [1.591]	0.0736*** [0.027]	0.1781*** [0.052]	0.0413 [0.041]	0.0311 [0.035]	0.0441 [0.029]
Female	3.4430*** [1.180]	4.3276*** [1.269]	4.1916*** [1.326]	2.8178** [1.380]	2.4348* [1.395]	0.0654*** [0.025]	0.1470*** [0.050]	0.0415 [0.036]	0.048 [0.032]	0.0249 [0.027]
FemalexTreatment	-5.6691*** [1.811]	-6.9564*** [1.967]	-6.0201*** [2.024]	-4.7690** [2.046]	-4.9307** [1.990]	-0.0858*** [0.031]	-0.1838*** [0.058]	-0.0513 [0.047]	-0.0471 [0.040]	-0.0609* [0.035]
Period	2.0356*** [0.123]					0.0145*** [0.004]				
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	640	160	160	160	160	640	160	160	160	160
Number of subject	160					160				
R-squared		0.156	0.115	0.083	0.061		0.153	0.031	0.036	0.043

Notes: Robust standard errors at the session level shown in brackets. * denotes significance at the 10% level, ** denotes significance at the 5% level and *** denotes significance at the 1% level. For a description of “Control Variables,” see notes of Table 2.

Table 4: Feedback Content on the Treatment Effect on Performance under Piece-Rate

PANEL A	Correct			Quality		
	PERIODS 2-4	PERIODS 3-4	PERIOD 4	PERIODS 2-4	PERIODS 3-4	PERIOD 4
Constant	10.4632*** [2.717]	13.5234*** [2.887]	10.4381*** [3.552]	0.8193*** [0.066]	0.8286*** [0.080]	0.7502*** [0.070]
Treatment	1.6195** [0.649]	1.3172* [0.681]	1.4808 [1.185]	0.0084 [0.027]	0.0022 [0.029]	0.0152 [0.023]
Positive	8.7330*** [0.703]	9.1939*** [0.716]	9.6846*** [1.040]	0.1115*** [0.020]	0.1177*** [0.022]	0.1323*** [0.021]
PositivexTreatment	1.0141 [1.057]	0.9917 [1.055]	0.5545 [1.452]	0.006 [0.028]	0.0079 [0.030]	-0.0091 [0.029]
Period	1.0387*** [0.164]	-0.1205 [0.290]		-0.0029 [0.005]	-0.0099 [0.007]	
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes
Observations	480	320	160	480	320	160
PANEL B	Correct (MALE)			Quality (MALE)		
Constant	11.3854*** [3.986]	11.5650*** [4.288]	8.0997 [5.794]	0.7835*** [0.121]	0.7170*** [0.135]	0.6607*** [0.124]
Treatment	1.9125* [1.134]	1.286 [1.326]	1.3685 [2.053]	0.0268 [0.043]	0.0341 [0.052]	0.0364 [0.044]
Positive	8.2538*** [1.251]	8.7481*** [1.300]	8.7700*** [1.752]	0.1213*** [0.036]	0.1618*** [0.036]	0.1534*** [0.038]
PositivexTreatment	2.9366 [1.948]	2.9605 [1.928]	2.5089 [2.459]	-0.0037 [0.048]	-0.0316 [0.052]	-0.0269 [0.053]
Period	0.8257** [0.320]	-0.0104 [0.469]		-0.0047 [0.009]	0.0018 [0.013]	
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes
Observations	180	120	60	180	120	60
PANEL C	Correct (FEMALE)			Quality (FEMALE)		
Constant	9.4470*** [3.295]	14.7880*** [3.492]	12.3310*** [4.599]	0.8510*** [0.081]	0.8858*** [0.094]	0.7825*** [0.087]
Treatment	1.3092* [0.753]	1.2006 [0.758]	1.5777 [1.518]	-0.0029 [0.035]	-0.022 [0.034]	-0.0025 [0.029]
Positive	8.4789*** [0.808]	9.1797*** [0.850]	10.0729*** [1.366]	0.1004*** [0.026]	0.0890*** [0.028]	0.1190*** [0.026]
PositivexTreatment	0.0161 [1.260]	-0.1558 [1.313]	-0.709 [1.861]	0.0121 [0.036]	0.0327 [0.036]	0.0035 [0.035]
Period	1.1700*** [0.192]	-0.1879 [0.366]		-0.0021 [0.006]	-0.0167* [0.009]	
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes
Observations	300	200	100	300	200	100

Notes: Robust standard errors at the session and subject level shown in brackets. * denotes significance at the 10% level, ** denotes significance at the 5% level and *** denotes significance at the 1% level. The variable *Positive* refers to being above average or improving performance. For a description of “Control Variables,” see notes of Table 2.

Table 5: Treatment Effect on Satisfaction under Piece-Rate: Happiness, Arousal, Dominance

	HAPPINESS					AROUSAL					DOMINANCE				
	ALL	PERIOD 1	PERIOD 2	PERIOD 3	PERIOD 4	ALL	PERIOD 1	PERIOD 2	PERIOD 3	PERIOD 4	ALL	PERIOD 1	PERIOD 2	PERIOD 3	PERIOD 4
Constant	5.5302*** [0.983]	6.3570*** [1.102]	5.6147*** [1.238]	5.3794*** [1.380]	5.2132*** [1.207]	8.7344*** [1.466]	10.5267*** [1.461]	8.6353*** [1.729]	8.7126*** [1.955]	9.5628*** [1.852]	5.5028*** [1.194]	5.0378*** [1.391]	5.8603*** [1.335]	5.6572*** [1.366]	5.3604*** [1.293]
Treatment	-0.2897 [0.207]	-0.1893 [0.248]	-0.3586 [0.261]	-0.4159 [0.268]	-0.187 [0.255]	0.0947 [0.238]	-0.0211 [0.268]	-0.0148 [0.266]	0.0633 [0.303]	0.3579 [0.325]	-0.1763 [0.223]	0.0195 [0.277]	-0.2512 [0.248]	-0.2752 [0.253]	-0.1287 [0.254]
Period	-0.034 [0.050]					-0.2599*** [0.063]					-0.03 [0.044]				
Female	0.2007 [0.200]	0.5875** [0.244]	0.0926 [0.253]	0.2108 [0.262]	-0.0773 [0.262]	-0.191 [0.266]	0.1721 [0.292]	0.3916 [0.306]	-0.0338 [0.339]	0.2298 [0.333]	-0.1908 [0.236]	0.067 [0.293]	-0.2156 [0.258]	-0.2547 [0.268]	-0.3264 [0.254]
Foreign	0.0486 [0.330]	-0.1639 [0.385]	0.138 [0.362]	0.0334 [0.462]	0.2022 [0.405]	-0.0037 [0.551]	0.295 [0.573]	0.2022 [0.582]	-0.3551 [0.646]	-0.16 [0.822]	0.0805 [0.395]	-0.18 [0.463]	-0.0826 [0.398]	0.3046 [0.469]	0.3196 [0.502]
Age	0.0161 [0.044]	0.0553 [0.046]	0.02 [0.049]	-0.0117 [0.061]	0.0026 [0.055]	0.1190* [0.069]	0.1839*** [0.067]	0.0981 [0.076]	0.081 [0.084]	0.1189 [0.088]	-0.0058 [0.054]	0.0198 [0.060]	-0.0044 [0.060]	-0.0279 [0.059]	-0.0129 [0.060]
Science	0.1547 [0.197]	0.2481 [0.237]	0.329 [0.249]	-0.1411 [0.258]	0.1814 [0.269]	0.0587 [0.250]	0.4476 [0.286]	-0.1068 [0.292]	-0.2284 [0.317]	0.1203 [0.332]	-0.1032 [0.234]	-0.2253 [0.280]	0.0296 [0.256]	-0.2544 [0.268]	0.0422 [0.257]
Math_Test	0.0384 [0.301]	-0.2458 [0.367]	0.2138 [0.390]	0.0986 [0.343]	-0.1022 [0.358]	0.5155 [0.315]	0.3357 [0.385]	0.6609 [0.400]	0.532 [0.440]	0.5471 [0.429]	0.2546 [0.303]	0.303 [0.400]	0.1298 [0.337]	0.4248 [0.359]	0.2291 [0.355]
Math_Test_Mis.	-0.4437 [0.552]	0.1449 [0.636]	-0.4645 [0.754]	-0.7035 [0.753]	-0.779 [0.665]	-0.6877 [0.727]	-1.1257 [0.760]	-1.2248 [0.763]	-0.1986 [1.066]	-0.2084 [0.922]	-0.542 [0.771]	-0.5064 [0.818]	-0.3717 [0.832]	-0.6329 [0.896]	-0.7846 [0.838]
Correct	0.0999*** [0.013]	0.0957*** [0.019]	0.0877*** [0.018]	0.1154*** [0.021]	0.1004*** [0.018]	0.0690*** [0.016]	0.0755*** [0.022]	0.0594** [0.024]	0.0890*** [0.026]	0.0422 [0.029]	0.0571*** [0.012]	0.0357* [0.018]	0.0385** [0.018]	0.0656*** [0.019]	0.0671*** [0.018]
Obs.	640	160	160	160	160	640	160	160	160	160	640	160	160	160	160
No. of subject	160					160					160				
R-squared		0.183	0.131	0.193	0.156		0.156	0.098	0.104	0.062		0.03	0.037	0.091	0.091

Notes: Robust standard errors at the session level shown in brackets. * denotes significance at the 10% level, ** denotes significance at the 5% level and *** denotes significance at the 1% level. *Happiness* takes values 1 to 9, where 1 represents the least happy and 9 represents the most happy. *Arousal* takes values 1 to 9, where 1 represents least aroused and 9 represents the most aroused. *Dominance* takes values 1 to 9, where 1 represents least dominant and 9 represents the most dominant. *Correct* measures the number of correct summations. For a description of all other variables, see notes of Table 2.

Table 6: Gender Differences on the Treatment Effect on Satisfaction under Piece-Rate: Happiness, Arousal, Dominance

	HAPPINESS					AROUSAL					DOMINANCE				
	ALL	PERIOD 1	PERIOD 2	PERIOD 3	PERIOD 4	ALL	PERIOD 1	PERIOD 2	PERIOD 3	PERIOD 4	ALL	PERIOD 1	PERIOD 2	PERIOD 3	PERIOD 4
Constant	5.5334*** [0.958]	6.2747*** [1.110]	5.6278*** [1.220]	5.4497*** [1.365]	5.2103*** [1.189]	8.7453*** [1.502]	10.4526*** [1.487]	8.5985*** [1.749]	8.8119*** [1.987]	9.6027*** [1.871]	5.4250*** [1.189]	5.0116*** [1.410]	5.8185*** [1.322]	5.6056*** [1.380]	5.2195*** [1.280]
Treatment	-0.2835 [0.336]	-0.409 [0.446]	-0.32 [0.447]	-0.1988 [0.493]	-0.1944 [0.423]	0.1159 [0.423]	-0.2188 [0.477]	-0.1235 [0.506]	0.3698 [0.544]	0.4609 [0.542]	-0.0326 [0.402]	0.0893 [0.558]	-0.1279 [0.472]	-0.116 [0.475]	0.2343 [0.438]
Female	0.2058 [0.286]	0.4122 [0.321]	0.1242 [0.386]	0.3877 [0.412]	-0.0833 [0.351]	0.2085 [0.393]	0.0144 [0.439]	0.3023 [0.469]	0.216 [0.527]	0.3132 [0.485]	0.0726 [0.387]	0.1227 [0.507]	-0.1143 [0.457]	-0.1249 [0.419]	-0.0321 [0.412]
FemalexTreat.	-0.0097 [0.420]	0.3324 [0.501]	-0.0598 [0.564]	-0.3389 [0.575]	0.0116 [0.519]	-0.0334 [0.507]	0.2992 [0.572]	0.1685 [0.589]	-0.4785 [0.651]	-0.161 [0.677]	-0.2269 [0.485]	-0.1056 [0.627]	-0.1912 [0.563]	-0.2486 [0.551]	-0.5677 [0.530]
Period	-0.0339 [0.050]					-0.2596*** [0.063]					-0.0287 [0.044]				
Correct	0.0999*** [0.013]	0.0993*** [0.020]	0.0872*** [0.020]	0.1129*** [0.022]	0.1005*** [0.018]	0.0689*** [0.016]	0.0787*** [0.022]	0.0610** [0.025]	0.0854*** [0.026]	0.041 [0.029]	0.0565*** [0.012]	0.0345* [0.020]	0.0367* [0.020]	0.0637*** [0.020]	0.0629*** [0.019]
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	640	160	160	160	160	640	160	160	160	160	640	160	160	160	160
No. of subject	160					160					160				
R-squared		0.185	0.131	0.195	0.156		0.158	0.098	0.108	0.062		0.031	0.038	0.093	0.099

Notes: Robust standard errors at the session level shown in brackets. * denotes significance at the 10% level, ** denotes significance at the 5% level and *** denotes significance at the 1% level. *Happiness* takes values 1 to 9, where 1 represents the least happy and 9 represents the most happy. *Arousal* takes values 1 to 9, where 1 represents least aroused and 9 represents the most aroused. *Dominance* takes values 1 to 9, where 1 represents least dominant and 9 represents the most dominant. For a description of “Control Variables,” see notes of Table 2.

Table 7: Feedback Content on the Treatment Effect on Satisfaction under Piece-Rate: Happiness, Arousal, Dominance

	HAPPINESS				AROUSAL				DOMINANCE			
	ALL	PERIOD 2-4	PERIOD 3-4	PERIOD 4	ALL	PERIOD 2-4	PERIOD 3-4	PERIOD 4	ALL	PERIOD 2-4	PERIOD 3-4	PERIOD 4
Constant	5.8266*** [0.969]	5.9528*** [1.068]	5.6189*** [1.154]	5.9565*** [1.232]	2.1738 [1.427]	2.8569* [1.602]	3.7865** [1.726]	1.2124 [1.746]	6.2882*** [1.238]	6.4984*** [1.232]	6.5623*** [1.274]	6.1740*** [1.257]
Treatment	-0.3126 [0.254]	-0.5838* [0.324]	-0.6745* [0.363]	-0.3735 [0.515]	0.1333 [0.250]	0.321 [0.351]	0.2185 [0.382]	-0.3301 [0.550]	-0.2083 [0.255]	-0.3857 [0.295]	-0.6186* [0.317]	-0.8315* [0.456]
Positive	0.3851** [0.164]	0.1577 [0.194]	0.4537** [0.187]	0.7087* [0.387]	0.3657** [0.184]	0.5721** [0.272]	0.3825 [0.271]	-0.2688 [0.490]	0.1777 [0.131]	0.1529 [0.206]	0.0688 [0.159]	0.2688 [0.407]
Positive x Treatment	0.4946** [0.246]	0.7301** [0.335]	0.9138*** [0.335]	0.5549 [0.580]	0.2322 [0.259]	-0.0679 [0.388]	0.1836 [0.418]	1.1379* [0.681]	0.3203* [0.187]	0.4474* [0.257]	0.8536*** [0.278]	1.2412** [0.543]
Period	0.0820* [0.045]	0.1603*** [0.060]	0.2380** [0.111]		-0.1866*** [0.057]	-0.3093*** [0.070]	-0.5293*** [0.126]		0.0398 [0.038]	0.0421 [0.039]	0.0378 [0.079]	
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	640	480	320	160	640	480	320	160	640	480	320	160
Number of subject	160	160	160		160	160	160		160	160	160	

Notes: Robust standard errors at the session level shown in brackets. * denotes significance at the 10% level, ** denotes significance at the 5% level and *** denotes significance at the 1% level. *Happiness* takes values 1 to 9, where 1 represents the least happy and 9 represents the most happy. *Arousal* takes values 1 to 9, where 1 represents least aroused and 9 represents the most aroused. *Dominance* takes values 1 to 9, where 1 represents least dominant and 9 represents the most dominant. The variable *Positive* refers to being above average or improving performance. For a description of “Control Variables,” see notes of Table 2.

Table 8: Treatment Effect on Performance under Flat-Rate

	Correct					Quality				
	ALL	PERIOD 1	PERIOD 2	PERIOD 3	PERIOD 4	ALL	PERIOD 1	PERIOD 2	PERIOD 3	PERIOD 4
Constant	-1.1144 [4.140]	5.0355 [4.087]	1.3366 [4.710]	2.7628 [4.628]	6.0615 [4.342]	0.6398*** [0.081]	0.6889*** [0.105]	0.5630*** [0.114]	0.6627*** [0.105]	0.7316*** [0.071]
Treatment	-0.0381 [0.946]	-0.3375 [0.954]	0.7478 [1.025]	-0.5104 [1.064]	-0.0522 [1.052]	-0.0128 [0.017]	0.0092 [0.025]	-0.0092 [0.023]	-0.0433** [0.022]	-0.0079 [0.018]
Period	1.9654*** [0.112]					0.0087** [0.004]				
Female	-1.3298 [1.007]	-1.3982 [1.012]	-1.4034 [1.110]	-0.9758 [1.138]	-1.5417 [1.104]	0.0048 [0.019]	-0.0051 [0.026]	0.015 [0.026]	0.0143 [0.024]	-0.0048 [0.018]
Foreign	1.2651 [1.457]	1.2808 [1.780]	0.6129 [1.567]	1.3183 [1.631]	1.8484 [1.579]	0.0432** [0.020]	0.0173 [0.026]	0.0818*** [0.027]	0.038 [0.026]	0.0358 [0.022]
Age	0.5566*** [0.173]	0.3880** [0.168]	0.6402*** [0.194]	0.6555*** [0.196]	0.5425*** [0.182]	0.0465** [0.021]	0.0781** [0.032]	0.0292 [0.030]	0.039 [0.025]	0.0398* [0.024]
Science	1.8667* [1.086]	1.4992 [1.036]	1.8632 [1.157]	1.9666 [1.247]	2.1377* [1.244]	0.0073** [0.003]	0.0056 [0.004]	0.0108** [0.004]	0.0080* [0.004]	0.0047 [0.003]
Math_Test	2.1871 [1.337]	0.9973 [1.300]	2.3826* [1.376]	2.6900* [1.525]	2.6785* [1.505]	0.0268 [0.026]	0.0209 [0.034]	0.0249 [0.034]	0.0304 [0.032]	0.0311 [0.027]
Observations	624	156	156	156	156		624	156	156	156
No. of subject	156					156				
R-squared		0.071	0.119	0.118	0.118		156	0.035	0.105	0.079

Notes: Robust standard errors at the session level shown in brackets. * denotes significance at the 10% level, ** denotes significance at the 5% level and *** denotes significance at the 1% level. For a description of all other variables, see notes of Table 2. Note that all subjects report whether they took the Math test.

Table 9: Gender Differences in the Treatment Effect on Performance under Flat-Rate										
	Correct					Quality				
	ALL	PERIOD 1	PERIOD 2	PERIOD 3	PERIOD 4	ALL	PERIOD 1	PERIOD 2	PERIOD 3	PERIOD 4
Constant	-1.7211 [4.353]	4.6744 [4.200]	0.7559 [4.938]	1.9064 [4.879]	5.4328 [4.600]	0.6338*** [0.083]	0.6882*** [0.107]	0.5569*** [0.115]	0.6546*** [0.109]	0.7222*** [0.073]
Treatment	2.3486 [1.503]	1.083 [1.509]	3.0321* [1.708]	2.8584* [1.700]	2.4208 [1.600]	0.011 [0.027]	0.0117 [0.039]	0.0148 [0.040]	-0.0117 [0.035]	0.0291 [0.024]
Female	0.6576 [1.361]	-0.2154 [1.329]	0.4987 [1.471]	1.8295 [1.519]	0.5176 [1.537]	0.0246 [0.024]	-0.003 [0.039]	0.0349 [0.033]	0.0406 [0.031]	0.026 [0.022]
FemalexTreatment	-4.0691** [1.930]	-2.4218 [1.939]	-3.8945* [2.171]	-5.7436*** [2.154]	-4.2163** [2.122]	-0.0405 [0.037]	-0.0043 [0.052]	-0.0408 [0.053]	-0.0539 [0.047]	-0.0631* [0.037]
Period	1.9654*** [0.112]					0.0087** [0.004]				
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	624	156	156	156	156	624	156	156	156	156
No. of subject	156					156				
R-squared		0.081	0.139	0.16	0.141		0.035	0.109	0.087	0.084

Notes: Robust standard errors at the session level are shown in brackets. * denotes significance at the 10% level, ** denotes significance at the 5% level and *** denotes significance at the 1% level. For a description of “Control Variables,” see notes of Table 2.

Table 10: Feedback Content on the Treatment Effect on Satisfaction under Flat-Rate: Happiness, Arousal, Dominance

	HAPPINESS				AROUSAL				DOMINANCE			
	ALL	PERIOD 2-4	PERIOD 3-4	PERIOD 4	ALL	PERIOD 2-4	PERIOD 3-4	PERIOD 4	ALL	PERIOD 2-4	PERIOD 3-4	PERIOD 4
Constant	6.4833*** [1.008]	5.9193*** [1.164]	5.8019*** [1.260]	6.2462*** [1.360]	4.7865*** [1.165]	4.2996*** [1.329]	3.8965*** [1.387]	2.4061* [1.358]	7.3387*** [0.914]	6.8193*** [0.943]	6.3606*** [0.935]	6.8641*** [1.198]
Treatment	0.0589 [0.268]	-0.2235 [0.355]	-0.1504 [0.381]	0.0484 [0.489]	-0.2035 [0.300]	-0.4216 [0.389]	-0.5102 [0.416]	0.0611 [0.488]	-0.2332 [0.254]	-0.3385 [0.278]	-0.2067 [0.295]	-0.4226 [0.431]
Positive	0.7005*** [0.161]	0.4328* [0.253]	0.8774*** [0.275]	1.1768*** [0.431]	0.1718 [0.238]	-0.1319 [0.282]	-0.2999 [0.343]	0.2208 [0.430]	0.3097** [0.129]	-0.0015 [0.144]	0.2753 [0.230]	0.1443 [0.379]
PositivexTreatment	-0.0937 [0.201]	0.2532 [0.349]	0.0344 [0.388]	-0.1521 [0.605]	-0.1062 [0.315]	0.1791 [0.409]	0.2912 [0.459]	-0.6221 [0.604]	0.0104 [0.173]	0.1796 [0.219]	-0.135 [0.277]	0.4683 [0.533]
Period	-0.0990** [0.048]	0.0178 [0.064]	0.0166 [0.100]		-0.0416 [0.056]	-0.0949 [0.059]	-0.1617 [0.100]		-0.0597* [0.033]	0.0123 [0.047]	0.1420* [0.076]	
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	624	468	312	156	624	468	312	156	624	468	312	156
No. of subject	156	156	156		156	156	156		156	156	156	

Notes: Robust standard errors at the subject and session level are shown in brackets. * denotes significance at the 10% level, ** denotes significance at the 5% level and *** denotes significance at the 1% level. For a description of “Control Variables,” see notes of Table 2.

Appendix

Instructions and the questionnaire were identical for the control and treated groups, except for the parts shown in bold, which appeared only in the treatment group.

A. Instructions:

THANK YOU FOR PARTICIPATING IN OUR EXPERIMENT!

This is an experiment and, thus, no talking, looking around or walking around is allowed. If you have any question or need help, please raise your hand and one of the researchers will assist you. If you do not comply with the rules, **WE WILL ASK YOU TO LEAVE THE EXPERIMENT AND YOU WILL NOT RECEIVE ANY PAYMENT.** Thank you.

This experiment is about individual decisions. Pompeu Fabra University has provided funds to carry it out. You will receive 3 euros for having arrived on time. Additionally, if you follow the instructions correctly, you may earn more money.

These instructions will inform you about the type of decisions you will be taking, as well as how your decisions will affect your payment. Everything you earn will be for you and paid in cash inside a closed envelope in a strictly private way at the end of the experimental session.

Each participant has been given an "Experiment Code" to guarantee that no participant can identify another one by his/her decisions or earnings. Researchers will observe each participant's earnings at the end of the experiment, but we will not associate your decisions with any participants' names.

Your Experiment Code is:

This experiment consists of four periods. Your final payment will be the sum of a participation fee of 3 euros plus whatever you earn in the four periods of the experiment.

Each period lasts 5 minutes. During this time you will be shown summations of four numbers of two digits each.

For example:

12
59
40
25

OK

The right solution is 136.

The summations will appear one by one and you will have to submit an answer in the indicated box. Using a calculator or paper and pencil for doing the summations is totally prohibited. If you do not comply with this rule, we will ask you to leave the experiment and you will not receive any payment. When you have solved a summation, you can submit the solution and click on "OK." The numbers in the summations, as well as the order in which they appear, will be exactly the same for all participants. In each period, you can solve as many summations as you can for the duration of the 5-minute time period.

You will be paid exactly 0.15 euros (15 cents) for each correct solution.

Thus, if you solve 1 summation correctly in the four periods, you will earn a total of 3.15 euros (3 euros as a show-up fee plus 0.15 euros for the correct solution).

Thus, if you solve 25 summations correctly in the four periods, you will earn a total of 6.75 euros (3 euros as a show-up fee plus $25 \times 0.15 = 3.75$ euros for the correct solutions).

Thus, if you solve 110 summations correctly in the four periods, you will earn a total of 19.5 euros (3 euros as a show-up fee plus $110 \times 0.15 = 16.5$ euros for the correct solutions).

Notice that the numbers in the examples are used for illustrative purposes. They DO NOT intend to suggest how many summations anyone should solve correctly.

Between the periods, you can rest for two minutes. During this time, you will be informed about the number of correctly solved summations during that period, **as well as about the average number of correctly solved summations in the experimental session**. Also, we will ask you to answer a brief questionnaire of three questions.

At the end of the 4 periods, you will be shown your total earnings for this experiment, **as well as the average earnings in this experimental session**, and we will ask you to fill in a questionnaire, as well as the information for the receipt. Wait for your Experiment Code to be called for you to come to the experimenter's room in order to receive the envelope with your earnings.

Thank you for your participation in our experiment!

B. Questionnaire:

A. Please, fill in the following information:

Gender
First Language
Field of Study
Year of Study
Age
Nationality

B. Questions:

1. Did you participate in similar experiments? If your answer is positive, please explain.
2. I am satisfied with the experience of having participated in this experiment.
 - a. In total disagreement
 - b. In disagreement
 - c. Neither in disagreement nor agreement
 - d. In agreement
 - e. In total agreement
3. I am satisfied with the payment that I obtained in this experiment.
 - a. In total disagreement
 - b. In disagreement
 - c. Neither in disagreement nor agreement
 - d. In agreement
 - e. In total agreement
4. I would consider participating again in this experiment.

- a. In total disagreement
- b. In disagreement
- c. Neither in disagreement nor agreement
- d. In agreement
- e. In total agreement

5. I value positively the information I obtained at the end of each period with respect to the number of summations I solved correctly.

- a. In total disagreement
- b. In disagreement
- c. Neither in disagreement nor agreement
- d. In agreement
- e. In total agreement

6. I value positively the information I obtained at the end of each period with respect to the average number of correct summations solved in this experimental session.

- a. In total disagreement
- b. In disagreement
- c. Neither in disagreement nor agreement
- d. In agreement
- e. In total agreement

7. Did you take the Math exam during *Selectividad*?

- a. Yes
- b. No

8. What grade did you obtain in your Math exam during *Selectividad*?