

The Provision of Relative Performance Feedback Information: An Experimental Analysis of Performance and Happiness*

Ghazala Azmat⁺

Nagore Iriberry^{**}

Universitat Pompeu Fabra and Barcelona GSE

April 2010

Abstract

This paper studies the effect of providing relative performance feedback information on individual performance and on individual affective response, when agents are rewarded according to their absolute performance. In a laboratory set-up, agents perform a real effort task and when receiving feedback, they are asked to rate their *happiness*, *arousal* and feeling of *dominance*. Control subjects learn only their absolute performance, while the treated subjects additionally learn the average performance in the session. Performance is 17 percent higher when relative performance feedback is provided. Furthermore, although feedback increases the performance independent of the content (i.e., performing above or below the average), the content is determinant for the affective response. When subjects are treated, the inequality in the happiness and the feeling of dominance between those subjects performing above and below the average increases by 8 and 6 percentage points, respectively.

Keywords: relative performance, piece-rate, feedback, social comparison, happiness.

JEL classification: I21, M52, C30.

* We thank Manel Baucells, Michael Bashshur, Gary Charness, George Loewenstein, Pedro Rey-Biel for their comments. Ghazala Azmat acknowledges financial support from ECO2008-06395-C05-01, Fundación Ramon Areces and the support of the Barcelona GSE Research Network and the Government of Catalonia. Nagore Iriberry acknowledges financial support from Fundación Rafael del Pino, Fundación Ramon Areces, Ministerio de Educación y Ciencia (ECO2009-11213 and SEJ2007-64340) and the support of the Barcelona GSE Research Network and the Government of Catalonia.

⁺ Ghazala Azmat. Departament d'Economia i Empresa. Universitat Pompeu Fabra, Ramón Trías Fargas 25-27, 08005 Barcelona (Spain). Tel: (+34) 935421757. E-mail: ghazala.azmat@upf.edu.

^{**} Nagore Iriberry. Departament d'Economia i Empresa. Universitat Pompeu Fabra, Ramón Trías Fargas 25-27, 08005 Barcelona (Spain). Tel: (+34) 935422690. E-mail: nagore.iriberri@upf.edu.

1. Introduction

Performance appraisals have become standard practice in organizations. As of the early 1980's, between seventy-four and eighty-nine percent of American businesses used them, see Murphy and Cleveland (1991). Informing agents about how well they are performing relative to their peers, in other words, by providing workers with relative performance feedback information, is a common way in which performance appraisals are implemented. Given its widespread use, it is important to understand the consequences of providing relative performance feedback information.

Managerial economics and social psychology has devoted quite a lot of attention to the study of performance appraisals, both to the theory and practice (see Bretz et al., 1992, and Levy et al., 2004, for reviews). Research on Economics, however, has paid little attention to relative performance feedback information. Two issues have, nevertheless, caught economists' attention. First, from a theoretical point of view the optimal provision of relative performance feedback information has been studied.¹ Second, empirical research has also focused on the effect of its provision on individual performance, which depends on the underlying incentives. In particular, under piece-rate incentives, when agents are rewarded according to their absolute performance, Hannan et al. (2008) in an experimental set-up without real effort, Azmat and Iriberry (2010) in an educational set-up, and Blanes i Vidal and Nossol (2009) in a firm setting, find that the provision of relative performance feedback information has a positive effect on performance. Eriksson et al. (2009), in an experimental set-up with real effort, find that although performance is not affected it does increase the mistake rate of the worst performing agent.^{2 3} A number of different reasons for why individuals would react to relative performance feedback information under piece-rate incentives have

¹ This work has mostly been conducted in the tournament literature, such as in Gershkov and Perry (2009), Kräkel (2007), Lai and Matros (2007) and Ederer (2010). Hansen (2009) studies the provision of performance appraisals when agents have career concerns, providing a rationale for the common finding that the provided feedback is less informative than it could be (mostly top-medium ratings and little negative feedback). Crutzen et al. (2010) studies the consequence of pay differentiation on employees' self-perception.

² Under flat-rate incentives, being observed and observing others' work brings positive effects on performance (Falk and Ichino, 2006, Mas and Moretti, 2009, and Kuhnen and Tymula, 2008). In tournaments, Casas-Arce and Martinez-Jerez (2009), Young et al. (1993), Muller and Schotter (2003), Hannan et al. (2008), Fehr and Ederer (2007) and Eriksson et al. (2009) study the effect of relative performance feedback information. The evidence is mixed, while some authors find that the provision of relative performance feedback increases all participants' effort, others find that the leading participants slack off and participants who are lagging behind give up. Ockenfels et al. (2010) study how bonus payments based on performance ratings, affect satisfaction and performance of managers in a large, multinational company.

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

been proposed. In particular, when relative performance feedback information is provided, the sense of competition among agents increases, which means they put in higher effort.

The papers mentioned above focus their studies on how agents' *performance* is affected when relative performance feedback information is provided. They do not pay any attention to how agents themselves *feel* when given relative performance feedback, despite its importance for practitioners and organizations. 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". One important measure of how agents feel is their level of happiness or subjective well-being. In recent years, economists and managers have started to give more weight to individuals' happiness as an important outcome variable of interest, along with more traditional outcome variables. Recent overviews of research on happiness and its relation to economics are provided by Kahneman et al. (1999), Loewenstein (2000), Frey and Stutzer (2002), Krueger (2005) and McFadden (2005). In particular, labor market studies on workers' psyche, which include broader measures than happiness, have shown to be very important in determining their moral and motivation, while at the same time having crucial effects on relevant issues such as firms' retention ability (see Bewley, 1999).

This paper studies the provision of relative performance feedback information under piece-rate incentives on two important measures: on agents' *performance* and on agents' *affective response*. Affective response includes measures of agents' emotional state, such as happiness (subjective well-being or experienced utility), arousal (or motivation) and feeling of dominance. Furthermore, given that in practice the relative performance feedback information is rarely provided just once, we investigate the *dynamic effects* of its provision on both, performance and affective response. Finally, by replicating our analysis under a flat-rate analysis scheme, we also investigate whether the *relevance* of the information, in terms of income, plays a role. By exploring all these dimensions, we provide a comprehensive analysis of the effects of the provision of relative performance feedback information, which, so far, have not been addressed.

We propose a controlled laboratory set-up, where subjects perform a real effort task and are rewarded according to piece-rate incentives. There are four working periods. In 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' affective response is elicited, that is, they are asked to rate their happiness, arousal and dominance levels. See Figure 1 for a graphical description of the experiment.

Our set-up offers a number of advantages over the previous studies for analyzing the effect of relative performance feedback information on performance under piece-rate incentives. In particular, when compared to field studies, the controlled lab environment allows us to attribute the observed treatment effect solely to the treated individuals' reaction and not to any other entity. For example, students who receive relative performance feedback information might be reacting through their parents' response, rather than their own (Azmat and Iriberry, 2010). Moreover, compared to the other two experimental papers, we use a real effort setting, unlike in Hannan et al. (2008), and the information provided refers to a group, unlike in Eriksson et al. (2009), who provide it as pair-wise comparison. In many labor settings, the comparison to a group of co-workers might be more informative than the comparison to one another co-worker. In addition to the set-up advantages, we are able to relate the performance findings to a number of new and interesting dimensions, such as gender, which have previously been unexplored.

With respect to performance, consistent with the mainstream of previous findings, we find that the provision of relative performance feedback information had a strong and positive effect on individual performance. Those subjects who received relative performance feedback information increased their performance by 17 percent compared to those who did not. With regard to the dynamic effects of providing relative performance feedback information, we find that in each period, the treated subjects outperform the untreated, although the effect becomes weaker over time. In addition, the *content* of the feedback information (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 increased their performance. Finally, we find a strong gender difference in the reaction to the treatment, such that solely boys drive the overall treatment effect. This is a new and interesting result that adds to the recent literature on gender differences in relevant economic environments.

With respect to the affective response, we find that the provision of relative performance feedback information had strong effects both on happiness and dominance

levels. Contrary to the findings on performance, we show that the treatment had very different effects on those who are receiving positive versus negative feedback. We find that receiving positive (negative) feedback affects subjects' happiness and dominance levels positively (negatively), such that when we only consider the overall treatment effect, the opposite signs cancel out. With respect to the happiness, the relative feedback leads to an increase in the gap (or inequality) of subjects' happiness, between those performing above and below the group average, by 8 percentage points. With respect to the dominance levels, the treatment leads to an increase in the inequality of subjects' feeling of dominance between those performing above and below the group average, by 6 percentage points. Moreover, the inequality in both happiness and dominance increases over time with the cumulative information. Finally, unlike the effect on performance, we find no gender differences in the affective response to the feedback treatment.

The finding related to the inequality in happiness is consistent with the finding on relative income and happiness, which has been studied in Economics since Duesenberry (1949) and Easterlin (1974) (see Clark et al., 2008, for a comprehensive and recent review). Those subjects who are made aware of the differences in income show a higher inequality in happiness than those subjects who are unaware of it. For further insight, we explore whether the effect we observe, both on performance and on affective response, are due to relative *performance* or relative *income* information. We investigate this by replicating our experiment under flat-rate incentives. Regarding individual performance, we find evidence that individuals mostly react when relative performance information has consequences in terms of income (i.e. under piece-rate but not under flat-rate incentives). Interestingly, we still find large gender difference in performance; while boys continue to react positively to the relative performance feedback, where the magnitudes are much smaller than under piece-rate, girls react negatively to it. In addition, we see no effect of relative performance feedback information on affective response under flat-rate incentives. These findings are important in two ways. First, relative performance feedback affects individuals' emotional state (increasing inequality in happiness and in the feeling of dominance) only when this information has consequences in terms of income (under piece-rate). Second, it also rules out a concern for potential experimental "demand" effects in the elicitation of subjects' affective response (i.e., subjects react to the information because they feel they are expected to by the experimentalist). Given the inequality in happiness

and feeling of dominance is only found under piece-rate incentives and not under flat-rate incentives, this effect is less likely to be driven by purely experimental demand effects.

To our knowledge this is the first paper that looks at the effect of relative performance feedback information on agents' affective response in a controlled environment. There are two related experimental papers. First, Charness and Grosskopf (2001) test in a dictator type setting 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 related to a preoccupation with relative payoffs.⁴ Note that our experimental design is rather different in terms of the underlying incentives. Unlike in their setting, the subjects in our environment can substantially improve their material payoff by putting in higher effort, while their effect on others' material payoffs is negligible. Second, Brandts et al. (2009) study the effect of a competitive environment on happiness and find that competition leads to an increase in inequality in happiness between those subjects who are chosen to participate and those who are excluded during the competition. This increase in inequality is consistent with our findings. Note, however, that in our setting the competition is rather symbolic, given that there is no actual prize or tournament-like incentives. However, the provision of relative performance feedback information might foster competition through social comparison (Festinger, 1954). See Azmat and Iriberry (2010) for further discussion on the relationship between relative performance feedback information and competition.

The findings from this paper have important implications for understanding whether or not an organization would choose to provide relative performance feedback information. We have shown that there are very strong effects on performance, when performance is rewarded, and this implies strong incentives for an organization to employ this mechanism, especially because its implementation has a negligible cost. However, we have gone beyond conventional thinking on this issue and we have highlighted that this mechanism, while being very effective on increasing performance, has important consequences on individuals' affective state. When agents' psychological aspects affect their morale and motivation, or even firms' retention ability, then it justifies why an organization would care about these aspects (Bewley, 1999).

⁴ The authors mention that this might be due to the fact that they find little concern for social standing and so their test has less power.

Organizations, only after understanding *all* the consequences of providing agents with relative performance feedback information can evaluate the appropriateness of such a policy.

The paper is organized as follows. Section 2 describes the experimental design and procedures in detail. Section 3 contains an overview of the main results, presented through graphical analysis. In Section 4 we proceed to do the econometric analysis, quantifying the main treatment effects. This section contains two main parts. First, we start by analyzing the effect of the treatment on individual performance. Second, we analyze the effect of the treatment on the individuals' affective response. Both subsections follow the same order: first we analyze the overall treatment effect, then, we test for gender differences and finally we analyze the effect of the positive/negative content of the feedback. Section 5 contains the analysis under a flat-rate incentive scheme. Finally, we conclude in Section 6.

2. Experimental Design and Procedures

Eight experimental sessions were conducted in the Laboratori d'Economia Experimental (LEEX) at Universitat Pompeu Fabra using z-Tree experimental software (Fischbacher, 2007) between April and May of 2009. A total of 160 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. After arrival, a sheet with general and identical instructions was distributed and read aloud to all subjects. The instructions can be found in the appendix at the end of the paper. Subjects were guaranteed a 3 euro show-up fee. Throughout the experiment we ensured effective separation between subjects, since they were seated at cubicles, so that they could not observe other subjects' screen. Once the experiment had concluded, subjects filled in a questionnaire while they waited to be paid. The questionnaire can be found in the appendix at the end of the paper. Subjects were paid individually and in private using a closed envelope. Each experimental session lasted one hour.

Figure 1 shows the timeline of the experiment. It 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 subjects were asked to submit an answer.⁵ Subjects were rewarded according to their

⁵ Many experimental papers have used summations, as a real-effort task, since it combines both ability and effort. See for example, Niederle and Vesterlund (2007) and Eriksson et al. (2009).

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 correctly solved summations during the four periods. We also replicated the whole experiment under flat-rate incentives (see Section 5). The only change we made to the experimental design was that, instead of rewarding subjects for every correct summation (piece-rate), we gave all subjects a flat-rate of 15 euros (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 information, as well as the SAM questionnaire.

The treatment variable of interest is the feedback information about performance, which was provided at the end of each period. Half of the subjects (80 subjects), the control group, were provided with information about their correct number of summations. We will refer to this as *absolute performance*. The other half (80 subjects), the treatment group, were provided not only with information about their correct number of summations, but they were also provided 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 the distance from this average. We will refer to this feedback information as the *relative performance* feedback information. It is important to note that all subjects, both in the treatment and control groups, were explained in the instructions the type of information they would receive (see instructions in the appendix at the end of the paper).

The SAM questionnaire, which is very common 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 in a numerical scale between 1 and 9, where 1 represents *happy* and 9 represents *unhappy*. The second set measures arousal in a scale between 1 and 9, where 1 represents *aroused* and 9 represents *unaroused*. The third set measures dominance in 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 affective response using the pictures provided. The SAM instructions included the list of words from the pertinent end of each semantic

differential scales in order to identify the anchors of each dimension to the subject. Thus, the subject was instructed, for example, to use the extreme happy SAM rating if the reaction was one of feeling “happy, pleased, satisfied, contented, hopeful, relaxed”, and to use the other extreme if she felt “unhappy, annoyed, unsatisfied, melancholic, despairing, or bored”. Similar instructions accompanied all three scales.⁶

Given that the SAM questionnaire was administered straight after providing subjects with feedback, we may be concerned that experimental demand effects might drive any effect found on individuals’ affective response. This means that subjects react to the feedback treatment because they think that the experimentalist expects them to react. That is a pervasive concern without a clean solution in any study that elicits subjects’ affective response. We address this in two ways. First, we follow a between-subject design methodology instead of a within-subject design, which is regarded as the safer methodology in order to avoid demand effects. Second, we conduct our experiment under both, piece-rate and flat-rate incentives, and find radically different results depending on the incentives. These differences, which are discussed at length in Section 5, rule-out this concern.

After the four working periods and the four periods of feedback and SAM questions are over, subjects were informed about their total earned money. The treatment group were also informed about the average earnings in the experiment, such that, again they could observe whether they would get higher or lower earnings than the average subject within the experimental session. The total average earnings in the experiment, including the show-up fee, were 14.69 euros, where the average was 14.55 euros in the control group and 14.84 euros in the treatment group.

Finally, while they waited to be paid, they were given a questionnaire that requested demographic details. We use these details as control variables in the econometric analysis. See the appendix for a specific list of questions they answered.

3. Graphical Analysis

⁶ The subject was instructed to use the extreme aroused SAM rating if the reaction was one of feeling “stimulated, excited, frenzied, jittery, wide awake, aroused” and to use the other extreme if she felt “relaxed, calm, sluggish, dull, sleepy, unaroused”. The subject was instructed to use the extreme dominated SAM rating if the reaction was one of feeling “controlled, influenced, cared for, awed, submissive”, and to use the other extreme if she felt “controlling, influential, in control, important, dominant”.

In this section, we provide an overall picture of the main results from the experiment using graphical analysis.

In Figures 3 to 6, we show subjects' *average* behavior for the treatment and control groups for the variables of interest. The variables of interest are the number of summations subjects submit (*Submitted*), which should be interpreted as effort, and the number of correct answers they give (*Correct*), which should be interpreted as performance. We are also interested in the happiness, arousal and dominance levels subjects reveal during the four periods. We do this in two steps. First, we look at averages over all subjects (Figures 3 and 4). Second, for a given period, we split the subject population between those who perform above the average in a given session, to which we will refer to as *above* subjects, and those who perform below the average, to whom we will refer to as *below* subjects (Figures 5 and 6).

Figure 3 shows the average submitted and correct number of summations over all subjects for each of the four periods. The treatment group submits significantly more summations than the control group in all four periods, although the significance declines over time (p -values of 0.00, 0.03, 0.04, and 0.07 for the two-sided mean comparison tests in periods 1 to 4, respectively). The treatment group also correctly solves significantly more summations than the control group in all four periods, although only marginally significant in the fourth period comparison (p -values of 0.00, 0.05, 0.06, and 0.10 for the two-sided mean comparison in periods 1 to 4, respectively).

The graphical analysis shows that the relative performance feedback information is an effective tool to increase performance. In addition, knowing that the relative performance feedback information will be provided is an important factor in pushing the individual performance, given the difference we observe in the first period. We also observe that there is learning over time for both the control and treatment groups. The average submitted and the average correctly solved summations increase over time, where learning is steeper in the second and third periods (see footnote 8).

Figure 4 shows the average emotional reaction using three main variables, happiness, arousal and dominance, respectively. As it can be seen in the graphs, the three variables show no significant difference between the treatment and the control groups (p -values for the difference between treatment and control, averaged over all 4 periods, is 0.76 for happiness, 0.40 for arousal and 0.71 for dominance). Note that in Figure 4 and from now on, we reverse the SAM scale for happiness and arousal to make

it more intuitive, such that 1 represents the least happy (unaroused) and 9 the most happy (aroused).

In Figure 5, the number of submitted, as well as correct, is higher in the treatment group for both, the above and below subjects, over all periods. The figures suggest that the feedback treatment increased performance for all subjects, independent of whether they were performing above or below the average. Also, it suggests that the treatment was not different for the below and above performing subjects, such that the gap between the above and below performing subjects remained the same for the control and treated subjects.

When we look at affective response separately for the above and below subjects, interesting patterns emerge. Regarding happiness, as one would expect, we can see from Figure 6 that above subjects are always happier than those below. This is the case for both the treatment and control groups. However, the inequality or gap in happiness between subjects performing above and below increases when subjects are provided with relative performance feedback information. This is an important finding because we observe that the provision of relative performance feedback information affects, not only subjects' performance, but also their happiness.

With respect to arousal, again we observe that in both, the control and treatment groups, subjects performing below the average show to be less aroused than subjects performing above the average. In addition, the relative performance feedback information seems to increase the level of arousal of those who are performing above. However, from the econometric analysis in the next section, we will see that this is not statistically significant.

Finally, regarding the feeling of dominance, again, as one would expect, in both the control and treatment groups, subjects performing above feel, overall, more dominant than subjects performing below the average. As in the case for happiness, the relative performance feedback information increases the gap in the feeling of dominance between those subjects performing above and below, compared to the control group.

To sum up, the treatment group shows higher performance than the control group. In addition, subjects performing above the average are overall happier, more motivated and more dominant than those subjects performing below the average. Furthermore, the provision of the relative performance feedback information increases the difference in happiness and in the feeling of dominance between those subjects who

are performing above the average and those who are performing below the average, compared to the control group. We now proceed to quantify these effects in the following section.

4. Econometric Analysis

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

4.1. The Effect of Relative Performance Feedback Information on Performance

4.1.1. The Overall Effect

The performance measures of interest are the number of summations subjects submit (*Submitted*) and the number of correct answers they get (*Correct*). Figure 7 shows the kernel distributions for these two variables. The distributions clearly show that for both, submitted and correct summations, 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 treated subjects, independent of the information they were receiving, that is, whether they were learning that they were performing above or below the average, which is in-line with those found in Azmat and Iriberry (2010).⁷ This will be confirmed in section 4.1.2.

To quantify the average treatment effect, we estimate the following linear regression.

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

where the dependent variable, Y_{it} , refers to the performance measures, *Submitted* or *Correct*, for individual i at period t . The variable *Treatment* identifies those who received the relative performance feedback information. We also include a time trend, *Period*.

Column 1 in Tables 1 and 2 show the estimates for equation (1) for submitted and correct summations, respectively. Since there are four working periods for each subject, we also weight the observations by the individual fixed effects (random effects

⁷ In Figure 2 of Azmat and Iriberry (2010) we can see similar kernel distributions.

model). On average, individuals submit 14.80 summations, of which 12.34 are correct. In both instances we see that the treated subjects significantly perform better. The treated subjects submit 13.82 percent (treatment coefficient 2.05) more summations and correctly answer 17.34 percent (treatment coefficient 2.14) more summations. It is interesting to note that the information, not only makes the subjects work harder but they also get better results. Time also matters and the estimates suggest that the subjects perform better over time. We further investigate the treatment effect separately for each period.

In columns 2 to 5 of Tables 1 and 2 we estimate by OLS the equation (1) for periods 1 to 4 for submitted and correct summations, respectively. Over time, subjects improve in their performance, which implies there is learning. Subjects become accustomed to the computer application and the task.⁸ We can also see that this learning is steeper in the beginning and it slows down in the last period. More interestingly, treated subjects outperform the control subjects in each period, although the effect is strongest in the initial period, suggesting an anticipation effect.⁹ We also observed that in Figure 3. The anticipation effect is in-line with those found by Blanes i Vidal and Nossol (2009) and Kuhnen and Tymula (2008), under piece-rate and flat-rate incentives, respectively.

An alternative performance variable of interest is the number of mistakes that subjects make during the task, i.e., the difference between the number of submitted summations and the number of correct summations. This would measure the quality of performance. Eriksson et al. (2009) have subjects that are in pairs performing under piece-rate incentives and each subject is informed about the performance of the other subject in her pair. They find that while the information had no effect per se on performance, it did increase the mistake rate of the worst performing subject. In order to test for the quality of performance, we estimated equation (1) using this outcome variable, but we find that the treatment had no effect on the quality of performance.

Finally, we repeat the analysis adding control variables.

$$(2) Y_{it} = \alpha + \beta Treatment_i + \lambda Period_t + \delta X_i + \varepsilon_{it}$$

⁸ Subjects were not allowed to use paper and pencil for summations. However, we observed that many subjects did the summations column by column. In the beginning, they were memorizing rather than putting into the computer the summation of the first column. By not having to memorize the summation for the first column, improved their performance substantially.

⁹ We observed the strong first period effect in all four sessions for the treated subjects.

where X_i is a vector of control variables, including gender, foreign, age and controls for ability. We use different measures of ability. First, we include a dummy for whether subjects' degree of study was science oriented or not (*Science_Degree*). Second, a dummy variable for whether the subject took the math test in the national level university entry test Selectividad (*Math_Test*).¹⁰ From columns 6-10 in Tables 1 and 2 we can see that the point estimates of the treatment effect overall, as well as period by period, are very similar to those without controls, as one would expect in a randomized experiment. As for controls, only the *Math_Test* variable has a small positive and significant effect in the number of submitted summations, although not in the number of correct summations.¹¹ Given that the estimates do not change when we include the control variables, in the analysis that follows we will estimate equation (1).

In summary, the provision of relative performance feedback information has a positive and significant effect on subjects' performance. This effect is strongest in the initial period. Furthermore, subjects continue to react to this information in all subsequent periods although the magnitude of the effect is lower than in the first period. The coefficients in periods 2, 3 and 4 are not significantly different from one another.¹²

4.1.2. Gender Analysis

Many studies have shown gender to be an important variable when looking at competitive environments (see Croson and Gneezy, 2009). Given that the provision of relative performance feedback information facilitates social comparison and that this might foster competition, we investigate whether girls react differently from boys to the relative performance feedback information.

We estimate the following equation:

$$(3) Y_{it} = \alpha + \beta Girl_i + \chi Treatment_i + \delta Girl_i * Treatment_i + \lambda Period_t + \varepsilon_{it}$$

¹⁰ 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 SAT in that it tests the knowledge on the topics covered during the last year of high school. If the student wants to do a science oriented degree she must take 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_Degree* and *Math_Test* is 0.27.

¹¹ We estimated the same regression with an alternative measure of ability which consists of the grade obtained in the Math test in Selectividad. We find quantitatively the same results. The sample size using the grade measure in Math test is reduced substantially since many subjects did not take this exam.

¹² When we combine all periods and compare pair-wise the treatment effect for each period, we find that they are not significantly different. The p -values for periods 2 and 3, periods 2 and 4, and periods 3 and 4 are 0.9830, 0.7700 and 0.7537, respectively.

where our variable of interest is the interaction between girls and the treatment. The results are shown in Table 3. We find very strong gender effects. Treated girls do not react differently from the non-treated girls, while treated boys do react positively compared to the non-treated boys. Notice that from our estimates in column 6, on average, a boy in the control group correctly solves 10.03 summations ($\hat{\alpha}$), while a treated boy solves 15.94 ($\hat{\alpha} + \hat{\chi}$) summations. However, on average, a girl in the control group correctly solves 13.52 ($\hat{\alpha} + \hat{\beta}$), while a treated girl correctly solves 13.45 ($\hat{\alpha} + \hat{\beta} + \hat{\chi} + \hat{\delta}$) summations. These results are striking since they imply that all of the observed effect on performance is attributed solely to boys. While treated boys increase their performance by 59%, treated girls do not change their performance at all.

Figure 8 shows the kernel distributions for the correct number of summations for boys and girls separately. These figures, again, clearly show that the effect is coming from boys reacting to the relative performance feedback information increasing their performance while girls do not react at all.

The strong gender effect in performance, when the relative performance feedback information is provided, is consistent with the empirical findings, mostly experimental, on both participation in competitions and on performance under competitive environments. 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 show to under-perform in competitive environments compared to men, mostly because men's performance increases when competing against women (Gneezy et al., 2003, Gneezy and Rustichini, 2004, and Antonovics et al., 2009, Hogarth et al., 2009). However, in a natural experiment on schooling, Azmat and Iriberry (2010) found that there is no gender differential effect to the provision of relative performance feedback information. Further research is needed to address the question of why gender differences might depend on the environment.

4.1.3. Feedback Content

We now consider the importance of the content of the feedback information. From period 2 until 4, subjects can react, not only to the provision of information, but also to the informational content. In other words, if a subject learns in a period that she

is performing above (below) the average, the content 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 that 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 that she has improved (worsened).¹³ 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 subjects who has performed above the average will not be informed that she is performing above. For simplicity, we refer to being above the average as positive feedback, irrespective of whether or not this information has been revealed.

We estimate the following regression:

$$(4) Y_{it} = \alpha + \beta Positive_{(t-1)i} + \chi Treatment_i + \delta Positive_{(t-1)i} * Treatment_i + \lambda Period_t + \varepsilon_{it}$$

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 (i.e., submitted and correct number summations, respectively). We cluster the standard errors at the subject level.

Panel A in Table 4 shows the estimates for equation (4) for the different periods. In columns 1 and 4, consistent with our previous analysis, all treated subjects outperform (both in terms of submitted (column 1) and correct number of summations (column 4), respectively) the 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 subject is given positive (negative) feedback, performance will increase. Estimations for the different

¹³ 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.

levels of cumulative feedback, columns 2-3 and 5-6, imply similar results. The results are however, somehow weaker given the limited number of observations as we approach the final period.

Since we found such a strong gender effect to the treatment, it is interesting to understand if boys and girls also react differently to the content of the information. In Panel B, we show the estimates for equation (4) for boys only, while in Panel C we show the estimates for girls only.

From Panel B, we see that boys react to the informational treatment, irrespective of whether they receive positive or negative feedback. This is consistent with the overall result. In Panel C, when we look at girls we find that the treatment, as well as the interaction of the content with treatment are not significant. This is consistent with our previous finding that girls do not react to either the informational treatment or the content of the feedback. Columns 2-3 and 5-6 give the estimates for the cumulative information. For boys, the estimates suggest the same findings but they become less significant due to fewer observations. Interestingly, for girls, although not significant, there are some differential effects depending on the content of the information. When girls receive negative feedback, the positive coefficients on the treatment variable, suggest that they work harder. However, when girls receive positive feedback the sign of the coefficient is negative, suggesting that they relax.

Overall, we find three important results. First, the provision of relative performance feedback information 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 information (positive or negative) does not affect subjects' subsequent performance differently. Third, we find a strong gender difference in the reaction to the treatment. The overall effect is driven solely by boys.

4.2. Affective Responses: Happiness, Arousal and Dominance.

4.2.1. The Overall Effect

Relative performance feedback information potentially has an effect on aspects other than performance. In particular, a subject's affective response may be influenced. Organizations care about the affective state of their employees, since it has been found that this affects productivity (see for example Iaffaldano and Muchinsky, 1985, Warr,

1999, and Oswald et al., 2009) and other relevant issues, such as workers' morale and firms' retention ability (Bewley, 1999). We measure three aspects of affective response: happiness (or subjective well-being), feelings of arousal and dominance.

We start by looking at the overall treatment effect on these measures. Equations (1) and (2) of section 4.1 are estimated, where the dependent variable, Y_{it} , now refers 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*.¹⁴ It is reasonable to assume that performance will have a direct effect on the affective response (see Frey and Stutzer, 2002). The results for the three variables of interest are shown in Table 5. The first column for each panel refers to the overall effect, controlling for time fixed effects and by 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 affective response, implying that the provision of relative performance feedback information is not affecting the subjects' well-being, arousal and/or feeling of dominance. As expected, the number of correct summations has a positive impact on all three affective response measures. The better the subjects perform, the happier, the more aroused and dominant they feel. The coefficient on *Correct* is positive and significant for all three measures of affective response. As for the other control variables in Table 5b, the only noticeable finding is that, as one would expect, subjects find the task less interesting over time (arousal level decreases). As with the performance analysis, since the control variables do not change the treatment coefficient, the subsequent analysis does not include them.

4.2.2. Gender

In section 4.1.2 we found that the provision of relative performance feedback information affected the performance of boys very differently from girls'. In particular, we found that the effect on performance was driven solely by the boys' reaction to the treatment. It is therefore, natural to investigate whether there are gender differences in the affective response.

¹⁴ We also estimated (1) and (2) without the variable *Correct* and the main results remain the same.

We extend the analysis from the previous section, in line with equation (3), to look for gender differences in the treatment effect. Table 6 shows the results. The treatment is insignificant. Additionally, the interaction coefficient of girl and the treatment implies that the provision of relative performance feedback information does not affect boys' affective response differently from girls'.

4.2.3. Feedback Content

As with performance, we now turn to study the informational content on the subject's affective response. We will estimate the following equation:

$$(5) Y_{it} = \alpha + \beta Positive_{it} + \chi Treatment_i + \delta Positive_{it} * Treatment_i + \lambda Period_t + \varepsilon_{it}$$

where Y_{it} refers to the affective response measures and $Positive_{it}$ is a dummy variable if the subject performed above the average or improved over time (as explained in section 4.1.3) and 0 otherwise. Note that unlike equation (4), the feedback content in the current period will affect the dependent variable in the same period (see timeline of the experiment in Figure 1). This information is provided to the treated subjects just before eliciting their affective response. As with the performance analysis, we also consider how the cumulative feedback affects affective response.

Table 7 shows the estimation for the differential effects of the content of the information on the affective response. 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 those subjects who are receiving positive and negative feedback. We find that a subject who is in the control group and performing below the average reports a happiness of 6.26 ($\hat{\alpha}$), while a control subject performing above the average reports a happiness of 6.66 ($\hat{\alpha} + \hat{\delta}$). However, a treated subject who is informed that she has performed below the average reports a happiness level of 5.95 ($\hat{\alpha} + \hat{\chi}$), while a treated subject who is informed that she has performed above the average reports an average happiness of 6.85 ($\hat{\alpha} + \hat{\beta} + \hat{\chi} + \hat{\delta}$). This implies that control subjects performing above the average are 7 percent happier than those performing below the average. For the treated subjects, those who are given positive feedback reveal 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 8 percentage points.

This result is a key finding, as it suggests that when we look at the average effect of the treatment on happiness, the effect is canceled out. However, once we control for the content of the information, we see that the treatment increases the difference in happiness between those who received positive and 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 while for the control group the gap in happiness between the above and below performing subjects remains stable (from 7 to 10 percent), for the treatment group, 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 below the average. We find that a subject who is in the control group and performing below the average reports a dominance level of 6.13 ($\hat{\alpha}$), while a control subject performing above the average reports a dominance level of 6.31 ($\hat{\alpha} + \hat{\delta}$). However, a treated subject who is informed that she has performed below the average reports a dominance level of 5.94 ($\hat{\alpha} + \hat{\chi}$), while a treated subject who is informed that she has performed above the average reports an average dominance of 6.45 ($\hat{\alpha} + \hat{\beta} + \hat{\chi} + \hat{\delta}$). This implies that control subjects performing above the average feel 3 percent more dominant than those performing below the average. For the treated subjects on the other hand, those who are given positive feedback reveal to be 9 percent more dominant than those who are given negative feedback. Overall, the feedback treatment leads to an increase in the gap of subjects' dominance by 6 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 25 percentage points. We find that while for the control group the gap in dominance between the above and below performing subjects remains stable (from 3 to 4 percent), for the treatment group, the gap increases substantially with the cumulative feedback (from 9 to 29 percent).

This section shows that the treatment has very different effects on the affective response of those who are receiving positive versus negative feedback. We found that receiving positive (negative) feedback affects subjects' happiness and dominance levels positively (negatively), such that when we only consider 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 subjects performing below average is of equal magnitude. Furthermore, the treatment increases the inequality in both happiness and dominance levels. Given that we elicit subjects' affective response after the feedback treatment, there is the concern that there may be experimental demand effects. We are able to rule out this concern in the following section.

5. Analysis under Flat-Rate Incentives

There are two important concerns that we may have with regard to the analysis so far. First, under piece-rate, we cannot disentangle whether the effects we observe are a consequence of relative *performance* or relative *payment* feedback. Second, with respect to the findings on affective response, we may be concerned that the treatment effects we found are driven by experimental demand effects. In this section, we address both of these issues by replicating our analysis under flat-rate incentives.

Under piece-rate incentives, given that the agents are rewarded according to their absolute performance, the feedback informs them about their relative *performance*, as well as their relative *payment* (or *income*). It is important to understand whether the individuals are reacting to either the relative performance or the relative payment 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 from the

income effect. We run experiments using exactly the same procedures and design, except for the fact that now all subjects were paid 15 euros, irrespective of their performance (flat-rate incentives).¹⁵

Overall, we find no effect of providing relative performance feedback on any of the two measures of performance, as shown in Table 8 and Figure 9. There is, however, an interesting gender difference, as observed under piece-rate incentives. As it can be seen in Table 9, there is evidence that boys react positively to the treatment, although the magnitudes are much smaller than under piece-rate incentives. The overall effects are significant for the number of submitted summations but not for the number of correctly submitted summations.

To quantify the gender differences under the different incentive schemes, recall that under piece-rate, boys increased their performance by 59%, while now we see from Table 9 that under flat-rate incentives, they increased by 19%, showing that although in the same direction, the effect is much smaller. Furthermore, we find that the interaction between treatment and girl is negative and significant, such that girls react negatively to the relative performance information. Recall that under piece-rate, girls did not increase their performance, while now we see that under flat-rate incentives, their performance was reduced by 16%. Figure 10 shows the kernel distributions for the correct number of summations for boys and girls separately. We can, again, clearly see that while treated boys increase their performance, treated girls reduce their performance. This explains why, when we pool boys and girls together, the overall effect under flat-rate is canceled out.

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 boys and girls react *positively* to the relative income information, girls and boys react in *opposite directions* to the purely relative performance information.

Regarding affective response, shown in Table 10, we find that the relative performance feedback information 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), we find

¹⁵ We had two sessions where we had 18 subjects instead of 20, yielding a total of 156 subjects instead of 160.

that treatment is not significant. This is reassuring as it strengthens the validity of our findings under piece-rate, since the results suggests that the effects on both happiness and feeling of dominance found under piece-rate incentives are a real response to *relevant* information and not driven by experimental demand effect. We refer to it as relevant information because the relative performance feedback information under piece-rate has consequences in terms of payment and income, while relative performance feedback information under flat-rate incentives does not. These differences in the responses help us to also rule out that the effects on affective response found under piece-rate incentives were due to purely experimental demand effects.

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

6. Conclusions

In this study we have shown that relative performance feedback information is an important tool to increase the performance of individuals, independent of the feedback content. Given the provision of this information is easy to implement and almost cost-free, it is an attractive policy to improve performance. However, we have also shown that the relative performance feedback information increases the inequality in individuals' happiness and feeling of dominance. This may imply a trade-off to a policy maker who is deciding whether or not to provide relative performance feedback information. Although it is agreed that enhancing performance is a positive result, the increase in inequality in affective response of individuals might be debatable and therefore, its evaluation is subject to the culture of the organization.

Further analysis suggests that individuals mostly react when relative performance information has consequences in terms of income (under piece-rate incentives). In practice, we mostly observe that relative performance feedback information is linked to relative income information. Even without a formal piece-rate scheme, performing above or below the average usually has consequences in terms of

income, such as the possibility of promotions/demotions or even the possibility of being fired. Thus, we believe that the relevant findings are those seen under piece-rate incentives.

A striking finding of the paper is the large gender difference in the effect on performance from the provision of relative performance feedback information. We find that, under piece-rate, boys and girls react very differently to the information. While boys react positively, such that their performance increases by 59%, girls do not react at all. This will have important policy implications, as it suggests that the effectiveness of the policy will largely depend on *who* is being treated. The findings from the flat-rate incentives also add to the interest on the gender difference issue. We find that treated boys react, even when there are no income consequences depending on performance, but they react much less (19%) than under piece-rate. Girls, on the other hand, react negatively when this information is provided under a flat-rate scheme. These gender differences can be reconciled. Under piece-rate incentives, relative performance feedback information includes both relative income and purely relative performance information. From the differences between flat and piece-rate incentives, we can infer that while boys react positively to both elements, girls react positively to the relative income information but negatively to purely relative performance information. Hence, under piece-rate, the effect on boys is larger than under flat-rate and the effect on girls is cancelled out. In terms of the affective response, we find that both boys' and girls' happiness is affected by the relative income information (i.e., under piece-rate but not flat-rate), which is consistent with the relative income and happiness literature.

This paper contributes to the growing interest in the study of the effects of providing relative performance feedback information in organizations. We propose a comprehensive study of providing relative performance feedback information, including performance measures, as well as variables that measure agents' emotional state. Moreover, the results of this paper open new research questions. In particular, long-run effects of the provision of relative performance feedback information need to be studied further, in particular using field studies.

References

Antonovics, K., Arcidiacono, P. and R. Walsh (2009) "The Effects of Gender Interactions in the Lab and in the Field," with Peter, *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", forthcoming in *Journal of Public Economics*.

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 (2009) "Tournaments without Prizes: Evidence from Personnel Records", Mimeo.

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.

Casas-Arce, P., and F. 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.

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.

Crosen, R. and U. Gneezy (2009) "Gender Differences in Preferences", *Journal of Economic Literature* 47(2), 1-27.

Crutzen, S. Y. B, Swank, O. H. and B. Visser (2010) "Confidence Management: on Interpersonal Comparisons in Teams", Mimeo.

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 (2009) "Tournament Incentives in the field: Gender Differences in the Workplace", Mimeo.
- Duesenberry, J.S. (1949) "Income, Saving and the Theory of Consumer Behaviour", Harvard University Press, Cambridge.
- Ederer, F. (2010) "Feedback and Motivation in Dynamic Tournaments", *Journal of Economics and Management Strategy*.
- 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*.
- 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 Pressure", *Journal of Labor Economics*, 24 (1), pp 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. 220 pp. \$24.95. ISBN 0-69 1-06998-0.
- Gershkov, A. and M. Perry (2009) "Tournaments with Midterm Reviews", *Games and Economic Behavior*, 66: 162-190.
- 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.

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.

Hansen, S. (2009) "What They Don't Know Can't Hurt You: The Benefits of Limited Feedback in Organizations", Mimeo.

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.

Kräkel, M. (2007) "Emotions in Tournaments", *Journal of Economic Behavior and Organization* 67, 204-214.

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 (2008) "Rank Expectations, Feedback and Social Hierarchies", Mimeo.

Lai, E. K. and A. Matros (2007) "Sequential Contests with Ability Revelation", Mimeo.

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, pp. 112-145.

Müller, W. and A. Schotter (2003) "Workaholics and Drop Outs in Optimal Organizations", Mimeo, New York University.

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
<http://www.nber.org/papers/w13922>.

Ockenfels, A., Sliwka, D., and P. Werner (2010) “Bonus Payments and Reference Point Violations”, IZA DP No. 4795.

Oswald A., Proto, E. and D. Sgroi (2009) “Happiness and Productivity”, IZA Discussion Paper No. 4645.

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

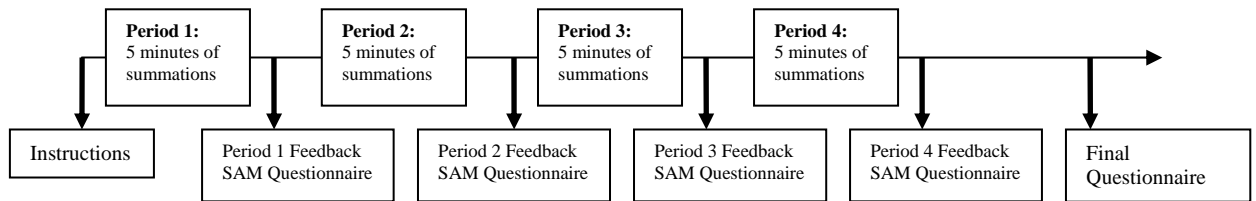
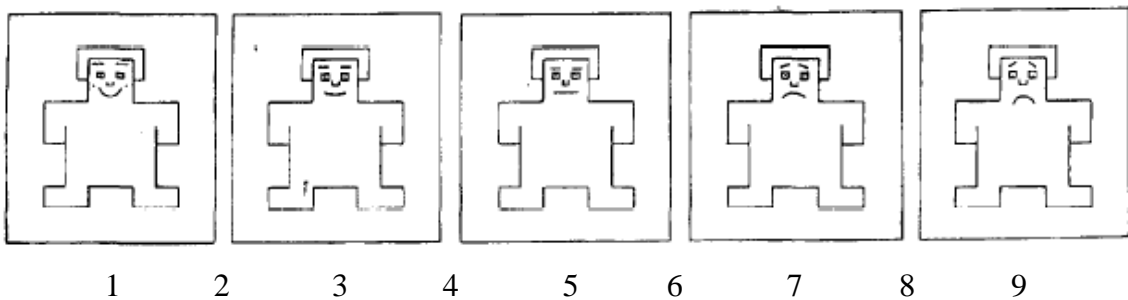


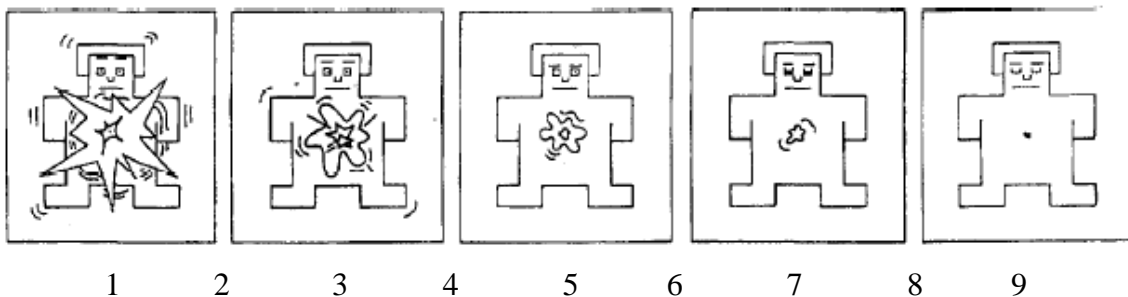
Figure 2. Self-Assessment Manikin (SAM)

Panel A: Happiness



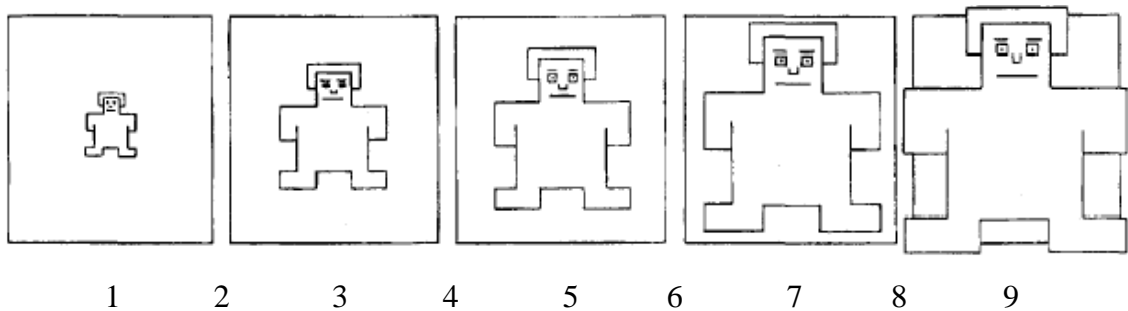
Note: 1 represents feeling “happy, pleased, satisfied, contented, hopeful, relaxed” and 9 represents feeling “unhappy, annoyed, unsatisfied, melancholic, despairing, or bored”.

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. Average Submitted and Correct Solutions

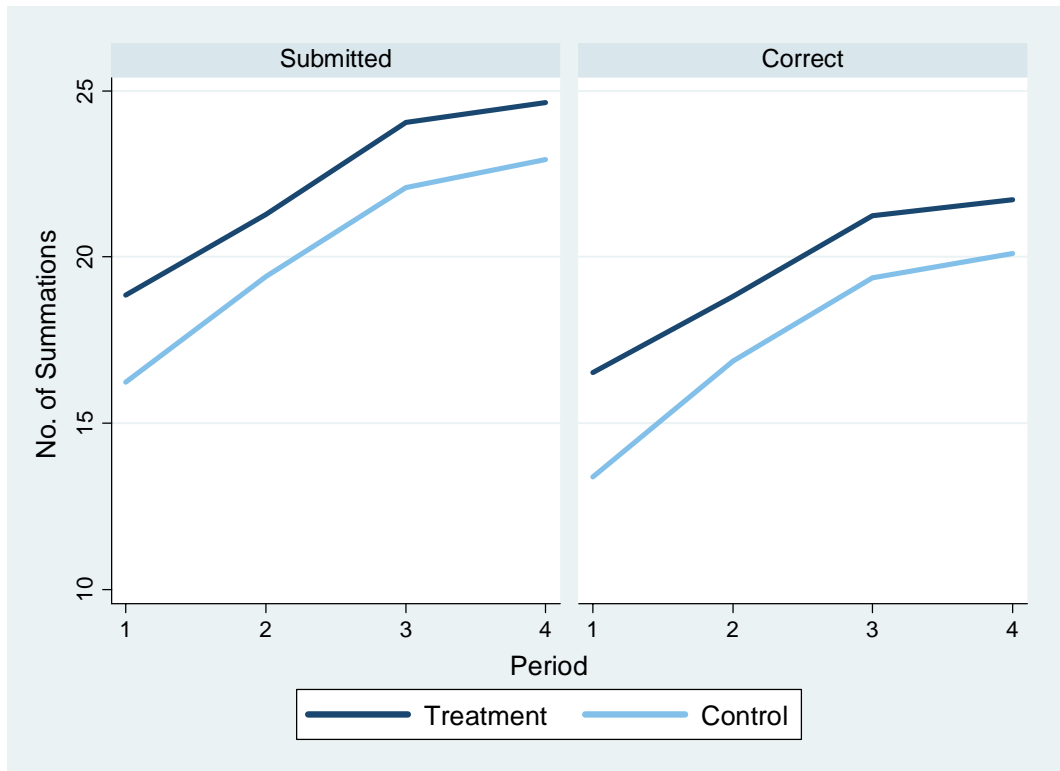
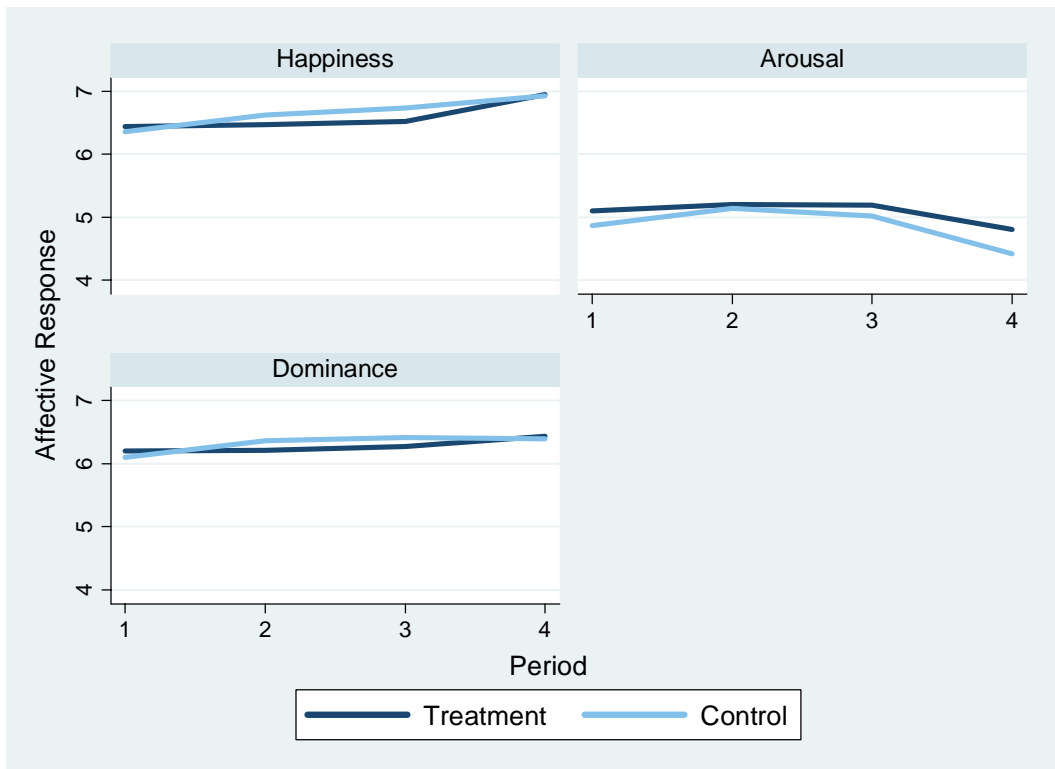


Figure 4. Average Happiness, Arousal and Dominance



Notes: *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.

Figure 5. Average Submitted and Correct Solutions for Above and Below

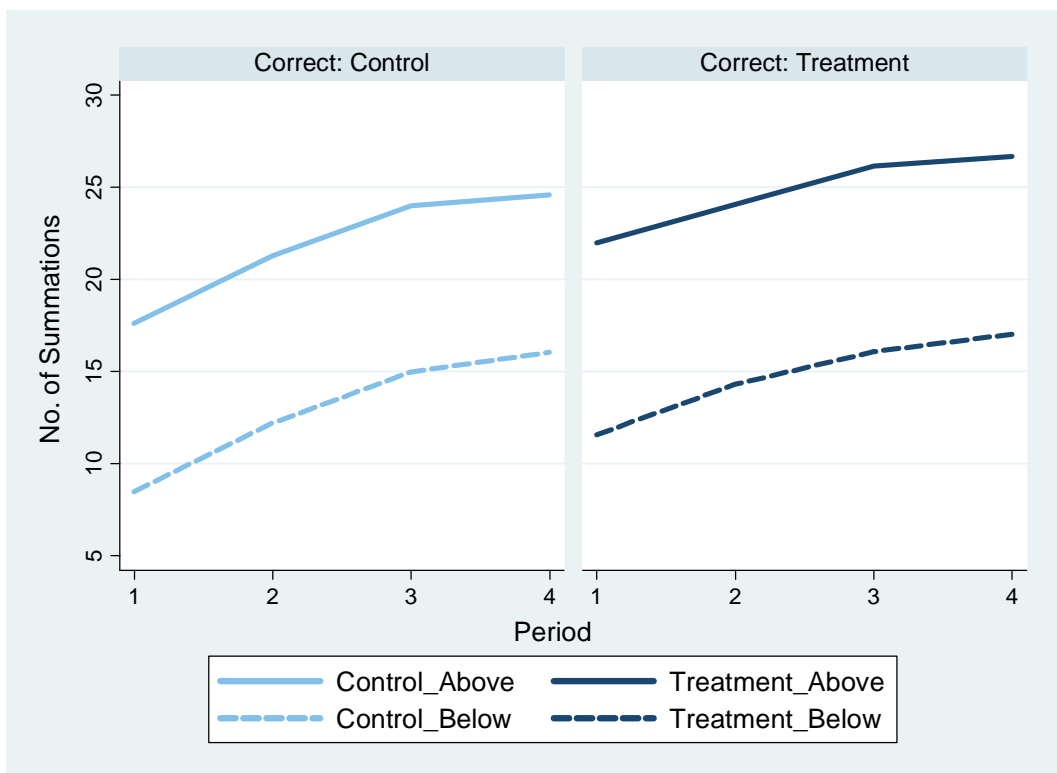
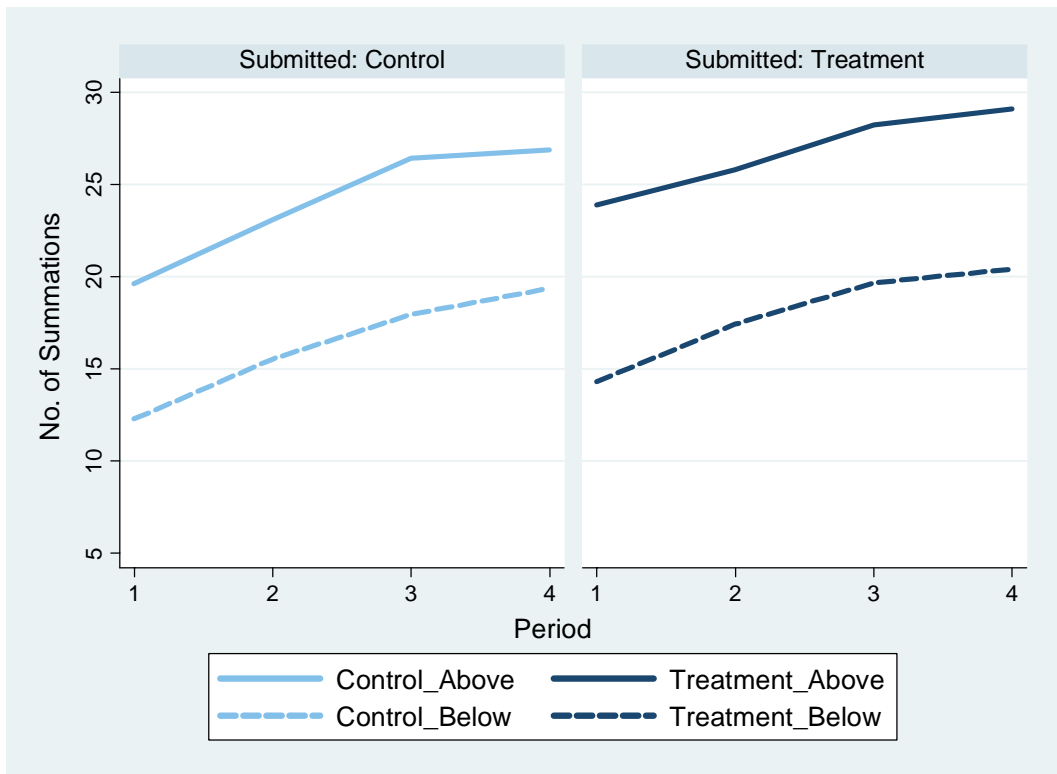
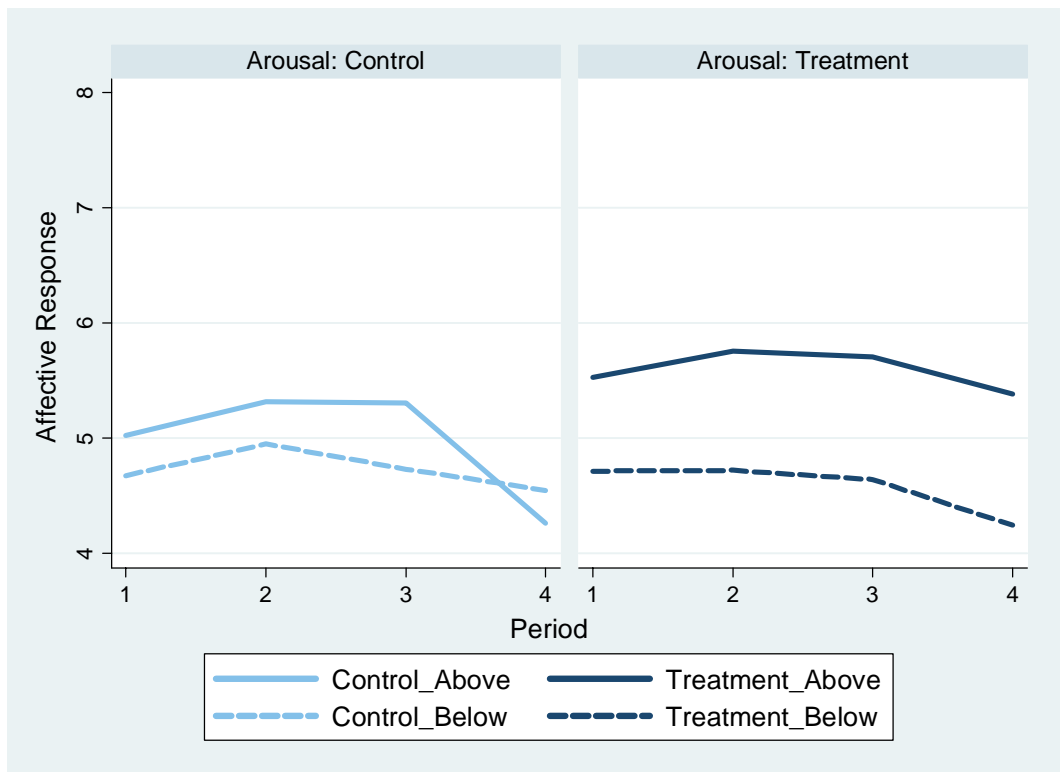
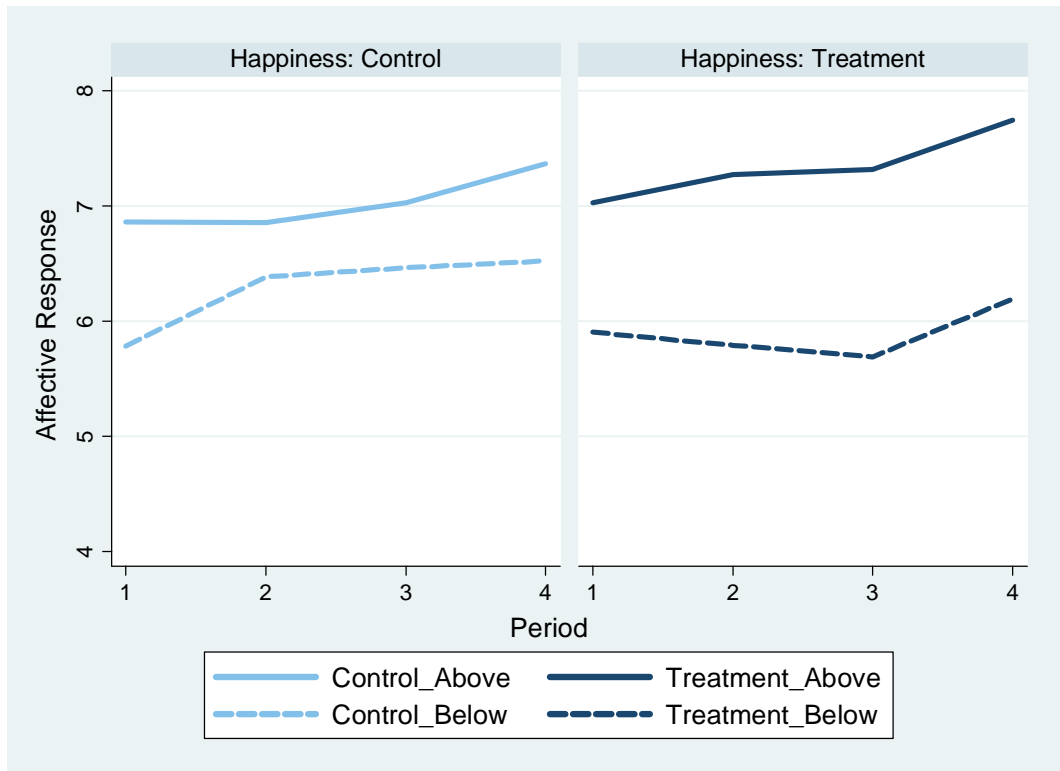
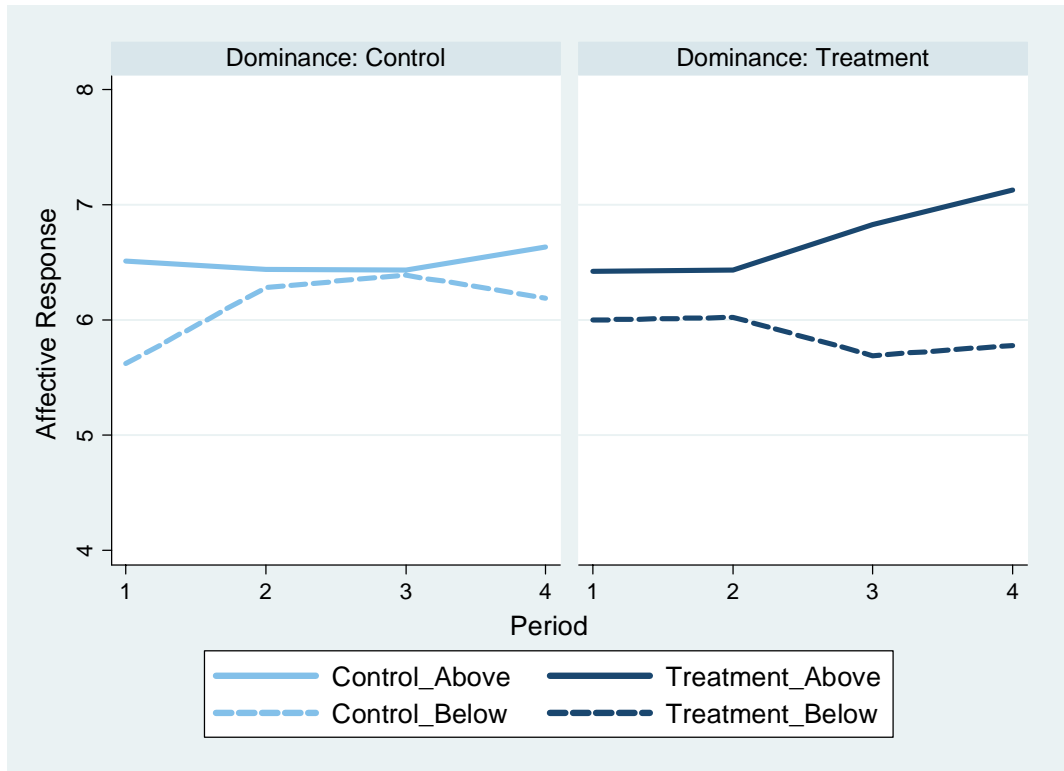


Figure 6. Average Happiness, Arousal and Dominance for Above and Below





Notes: *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.

Figure 7: Kernel Distributions

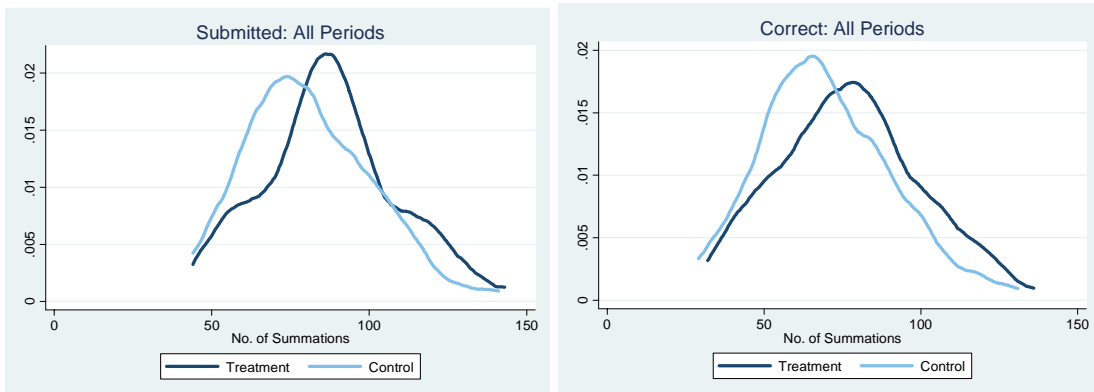


Figure 8: Kernel Distributions by Gender

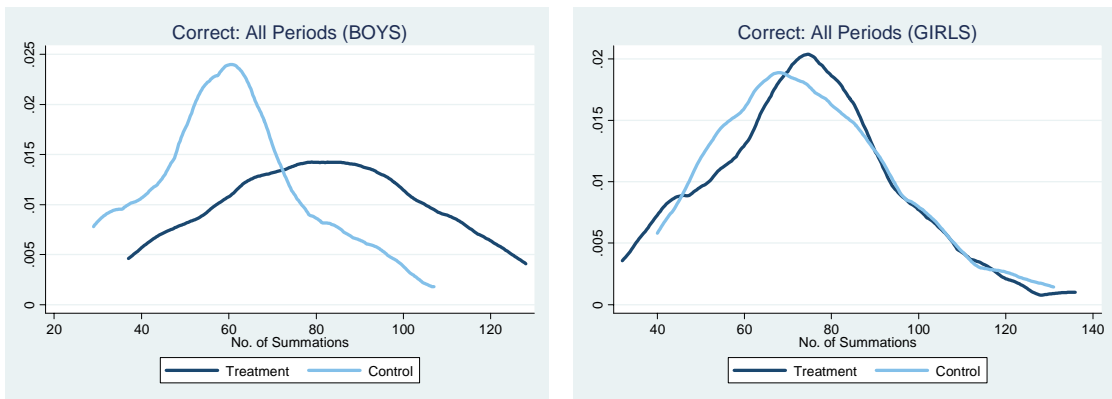


Figure 9: Kernel Distributions under Flat-Rate

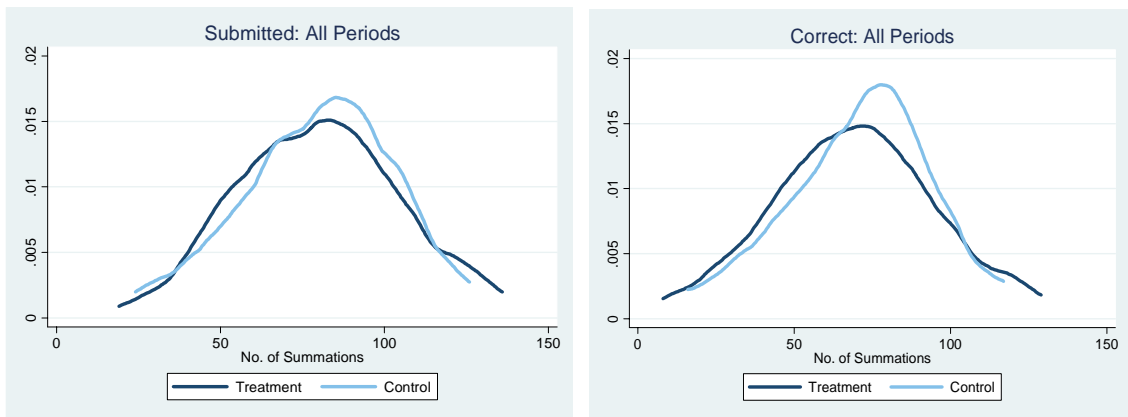


Figure 10: Kernel Distributions under Flat-Rate by Gender

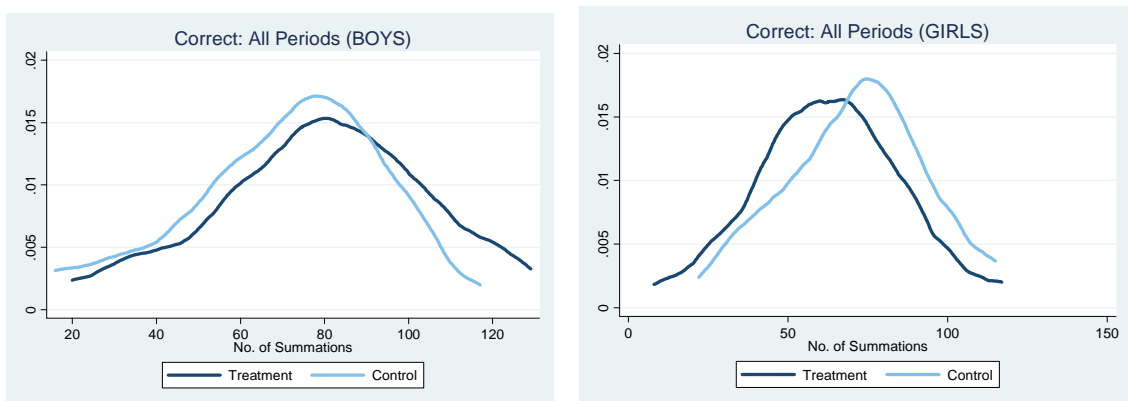


Table 1: Treatment Effect on Performance: Number of Submitted Summations

	ALL	PERIOD 1	PERIOD 2	PERIOD 3	PERIOD 4	ALL	PERIOD 1	PERIOD 2	PERIOD 3	PERIOD 4
Constant	14.8047*** [0.6487]	16.2375*** [0.6321]	19.4125*** [0.6530]	22.0875*** [0.6505]	22.9375*** [0.6723]	14.1366*** [4.9005]	16.6414*** [4.9700]	19.2984*** [5.1468]	22.9889*** [5.1619]	18.8939*** [5.3940]
Treatment	2.0469** [0.8711]	2.6250*** [0.8939]	1.8750** [0.9235]	1.9750** [0.9200]	1.7125* [0.9508]	2.0143** [0.9222]	2.5503*** [0.9361]	1.8141* [0.9694]	1.9538** [0.9722]	1.7391* [1.0159]
Period	2.1456*** [0.0815]					2.1276*** [0.0833]				
Girl						0.3718 [0.9569]	0.4096 [0.9714]	1.054 [1.0059]	0.0454 [1.0089]	-0.0218 [1.0542]
Foreign						-0.148 [2.0569]	-0.0811 [2.0880]	0.0169 [2.1623]	0.2709 [2.1686]	-0.7988 [2.2661]
Age						-0.0503 [0.2180]	-0.1112 [0.2213]	-0.1128 [0.2292]	-0.1135 [0.2299]	0.1364 [0.2402]
Science_Degree						0.1751 [0.9785]	0.0073 [0.9932]	0.4377 [1.0286]	-0.1891 [1.0316]	0.4444 [1.0780]
Math_Test						2.0677* [1.1971]	2.3797* [1.2152]	2.2962* [1.2584]	2.1243* [1.2621]	1.4706 [1.3189]
Observations	640	160	160	160	160	608	152	152	152	152
Number of subject	160					152				

Notes: * denotes significance at the 10% level, ** denotes significance at the 5% and *** denotes significance at the 1% level. The variable *Science_Degree* takes value 1 if the degree of study is Architecture, Biology, Business, Computer Science, Economics, Engineering or Human Resources and 0 if the degree is 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”.

Table 2: Treatment Effect on Performance: Number of Correct Summations										
	ALL	PERIOD 1	PERIOD 2	PERIOD 3	PERIOD 4	ALL	PERIOD 1	PERIOD 2	PERIOD 3	PERIOD 4
Constant	12.3422*** [0.6853]	13.3875*** [0.6909]	16.8625*** [0.6910]	19.3750*** [0.6867]	20.1000*** [0.6950]	11.5499** [5.1087]	12.1676** [5.4980]	16.3306*** [5.4754]	21.5645*** [5.4741]	16.3278*** [5.6040]
Treatment	2.1406** [0.9023]	3.1250*** [0.9770]	1.9500** [0.9772]	1.8625* [0.9712]	1.625 [0.9829]	2.1354** [0.9610]	3.1863*** [1.0355]	1.8396* [1.0313]	1.8368* [1.0310]	1.679 [1.0555]
Period	2.0356*** [0.1000]					2.0191*** [0.1034]				
Girl						0.5525 [0.9972]	0.7837 [1.0745]	1.237 [1.0701]	0.3527 [1.0699]	-0.1634 [1.0953]
Foreign						0.3086 [2.1435]	-0.1819 [2.3098]	0.9881 [2.3003]	0.6244 [2.2998]	-0.1962 [2.3543]
Age						-0.048 [0.2272]	-0.0433 [0.2448]	-0.102 [0.2438]	-0.1801 [0.2438]	0.1335 [0.2496]
Science_Degree						0.236 [1.0196]	0.0655 [1.0987]	0.4852 [1.0942]	0.3875 [1.0940]	0.0056 [1.1199]
Math_Test						1.8632 [1.2475]	2.1648 [1.3443]	2.2689* [1.3387]	1.5963 [1.3384]	1.4226 [1.3702]
Observations	640	160	160	160	160	608	152	152	152	152
Number of subject	160					152				

Notes: * denotes significance at the 10% level, ** denotes significance at the 5% and *** denotes significance at the 1% level. The variable *Science_Degree* takes value 1 if the degree of study is Architecture, Biology, Business, Computer Science, Economics, Engineering or Human Resources and 0 if the degree is 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”.

Table 3: Gender Differences in the Treatment Effect on Performance										
	Number of Submitted Summations					Number of Correct Summations				
	ALL	PERIOD 1	PERIOD 2	PERIOD 3	PERIOD 4	ALL	PERIOD 1	PERIOD 2	PERIOD 3	PERIOD 4
Constant	12.8859*** [1.0584]	14.2593*** [1.0657]	16.7407*** [1.0897]	20.4074*** [1.1036]	21.5926*** [1.1470]	10.0313*** [1.0973]	10.4815*** [1.1455]	14.0741*** [1.1558]	17.4444*** [1.1647]	18.4815*** [1.1799]
Treatment	5.3106*** [1.4005]	5.9832*** [1.4370]	5.8653*** [1.4693]	5.0168*** [1.4881]	4.3771*** [1.5466]	5.9099*** [1.4407]	7.7306*** [1.5445]	6.0168*** [1.5585]	5.1010*** [1.5704]	4.7912*** [1.5910]
Girl	2.8962** [1.2760]	2.9860** [1.3093]	4.0328*** [1.3388]	2.5360* [1.3559]	2.03 [1.4092]	3.4881*** [1.3127]	4.3864*** [1.4073]	4.2089*** [1.4200]	2.9140** [1.4309]	2.4430* [1.4496]
GirlxTreatment	-5.1856*** [1.7694]	-5.3348*** [1.8155]	-6.2772*** [1.8563]	-4.8538** [1.8800]	-4.2763** [1.9540]	-5.9706*** [1.8202]	-7.2794*** [1.9514]	-6.3850*** [1.9690]	-5.1404** [1.9840]	-5.0775** [2.0100]
Period	2.1456*** [0.0815]					2.0356*** [0.1000]				
Observations	640	160	160	160	160	640	160	160	160	160
No. of subject	160					160				

Notes: * denotes significance at the 10% level, ** denotes significance at the 5% and *** denotes significance at the 1% level.

Table 4: Feedback Content on the Treatment Effect on Performance

Table 4: Feedback Content on the Treatment Effect on Performance						
PANEL A	Submitted			Correct		
	PERIODS 2-4	PERIODS 3-4	PERIOD 4	PERIODS 2-4	PERIODS 3-4	PERIOD 4
Constant	12.5459*** [0.7131]	16.6226*** [1.3285]	19.1220*** [0.7139]	10.2386*** [0.7381]	14.0976*** [1.5413]	16.1220*** [0.7268]
Treatment	1.7487** [0.6737]	1.7596** [0.7159]	1.4422 [1.0224]	1.5774** [0.7166]	1.5832* [0.8089]	1.1601 [1.0409]
Positive	7.2548*** [0.7835]	7.4654*** [0.8320]	8.1508*** [1.0690]	7.5776*** [0.8110]	7.8124*** [0.8682]	8.2114*** [1.0883]
PositivexTreatment	0.656 [1.2872]	0.364 [1.3272]	0.2062 [1.4927]	0.9437 [1.2989]	0.5296 [1.3527]	0.796 [1.5197]
Period	1.7384*** [0.1717]	0.6163* [0.3524]		1.5523*** [0.1907]	0.4954 [0.4075]	
Observations	480	320	151	480	320	151
PANEL B	Submitted (BOYS)			Correct (BOYS)		
	PERIODS 2-4	PERIODS 3-4	PERIOD 4	PERIODS 2-4	PERIODS 3-4	PERIOD 4
Constant	11.5644*** [1.1121]	18.0263*** [2.0654]	19.0000*** [1.1172]	9.1398*** [1.1628]	15.1243*** [2.6368]	16.1111*** [1.0997]
Treatment	2.5860** [0.9996]	2.4603** [1.0350]	1.4615 [1.7252]	2.4212** [1.2002]	2.2524* [1.3442]	1.5043 [1.6982]
Positive	7.5287*** [1.2600]	8.3232*** [1.2584]	7.5714*** [2.1113]	7.5147*** [1.4220]	7.8620*** [1.4290]	6.4603*** [2.0783]
PositivexTreatment	1.4859 [2.1771]	0.5812 [2.1664]	1.8618 [2.7145]	2.1908 [2.2320]	1.6142 [2.2435]	3.1348 [2.6720]
Period	1.8664*** [0.3043]	0.101 [0.5903]		1.7049*** [0.3219]	0.1039 [0.7058]	
Observations	180	120	57	180	120	57
PANEL C	Submitted (GIRLS)			Correct (GIRLS)		
	PERIODS 2-4	PERIODS 3-4	PERIOD 4	PERIODS 2-4	PERIODS 3-4	PERIOD 4
Constant	13.3570*** [0.9358]	16.0972*** [1.7649]	19.2174*** [0.9427]	11.1313*** [0.9248]	13.7413*** [1.8771]	16.1304*** [0.9750]
Treatment	1.0783 [0.9020]	1.2366 [0.9744]	1.398 [1.2941]	0.8889 [0.8729]	1.0957 [0.9973]	0.9849 [1.3385]
Positive	6.8018*** [0.9707]	7.0221*** [1.0417]	8.2441*** [1.2941]	7.1864*** [0.9505]	7.6146*** [1.0277]	8.6773*** [1.3385]
PositivexTreatment	0.0705 [1.4590]	-0.154 [1.5356]	-0.9122 [1.8806]	0.2366 [1.5064]	-0.3023 [1.5973]	-0.4242 [1.9450]
Period	1.6467*** [0.2096]	0.8607* [0.4420]		1.4466*** [0.2392]	0.6784 [0.4978]	
Observations	300	200	94	300	200	94

Notes: * denotes significance at the 10% level, ** denotes significance at the 5% and *** denotes significance at the 1% level. The variable *Positive* becomes positive feedback for the treated subjects while this information is never revealed to the control subjects. We cluster the standard errors at the individual level.

Table 5a: Treatment Effect on Affective Response: 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	4.9975*** [0.2224]	5.0869*** [0.3012]	5.0800*** [0.3838]	4.4660*** [0.4260]	4.8887*** [0.4322]	4.2881*** [0.2636]	3.8347*** [0.3506]	4.0353*** [0.4117]	3.2930*** [0.5115]	3.4930*** [0.5616]	5.3919*** [0.2159]	5.5858*** [0.3404]	5.6999*** [0.3659]	5.0916*** [0.4238]	5.0224*** [0.4319]
Treatment	-0.281 [0.2021]	-0.2228 [0.2392]	-0.3287 [0.2517]	-0.4309* [0.2480]	-0.1396 [0.2457]	0.0653 [0.2416]	-0.0024 [0.2784]	-0.065 [0.2700]	0.0097 [0.2978]	0.3132 [0.3193]	-0.1609 [0.2236]	-0.02 [0.2704]	-0.2266 [0.2399]	-0.2645 [0.2467]	-0.0739 [0.2455]
Period	-0.0354 [0.0428]					-0.2623*** [0.0502]					-0.031 [0.0350]				
Correct	0.1006*** [0.0119]	0.0953*** [0.0189]	0.0916*** [0.0202]	0.1172*** [0.0201]	0.1013*** [0.0197]	0.0702*** [0.0141]	0.0768*** [0.0220]	0.0654*** [0.0217]	0.0888*** [0.0241]	0.0457* [0.0256]	0.0576*** [0.0106]	0.0384* [0.0213]	0.0393** [0.0193]	0.0682*** [0.0200]	0.0685*** [0.0197]
Observations	640	160	160	160	160	640	160	160	160	160	640	160	160	160	160
No. of subject	160					160					160				

Notes: * denotes significance at the 10% level, ** denotes significance at the 5% 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.

Table 5b: Treatment Effect on Affective Response: 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	4.7837*** [1.1384]	4.1427*** [1.3068]	4.9326*** [1.4329]	4.9001*** [1.4432]	4.7682*** [1.4057]	0.6965 [1.3399]	-1.1634 [1.4918]	0.9229 [1.5294]	0.2215 [1.7073]	0.1027 [1.8276]	5.6337*** [1.2389]	5.3264*** [1.4872]	5.9627*** [1.3576]	5.5936*** [1.4094]	5.4434*** [1.3837]
Treatment	-0.255 [0.2137]	-0.166 [0.2499]	-0.2866 [0.2648]	-0.3513 [0.2612]	-0.2186 [0.2596]	0.0803 [0.2515]	-0.09 [0.2852]	-0.0174 [0.2827]	0.0365 [0.3090]	0.3764 [0.3375]	-0.2322 [0.2329]	-0.0334 [0.2843]	-0.3184 [0.2509]	-0.3329 [0.2550]	-0.1947 [0.2555]
Period	-0.0206 [0.0438]					-0.2673*** [0.0516]					-0.0283 [0.0360]				
Girl	0.2751 [0.2202]	0.6550** [0.2517]	0.1563 [0.2731]	0.2948 [0.2682]	0.0058 [0.2670]	0.1486 [0.2591]	0.1225 [0.2873]	0.3333 [0.2915]	-0.0757 [0.3173]	0.2 [0.3472]	-0.1371 [0.2406]	0.0927 [0.2864]	-0.1635 [0.2587]	-0.1913 [0.2619]	-0.2523 [0.2629]
Foreign	-0.0235 [0.4730]	-0.3512 [0.5400]	0.0939 [0.5847]	-0.025 [0.5764]	0.1885 [0.5740]	0.2727 [0.5568]	0.5536 [0.6164]	0.5355 [0.6241]	0.0028 [0.6819]	-0.0122 [0.7463]	0.172 [0.5169]	0.0215 [0.6145]	-0.0197 [0.5540]	0.3785 [0.5629]	0.3176 [0.5650]
Age	-0.0012 [0.0501]	0.0328 [0.0572]	-0.0088 [0.0620]	-0.0271 [0.0612]	-0.0005 [0.0609]	0.1483** [0.0590]	0.2138*** [0.0653]	0.1160* [0.0661]	0.1331* [0.0724]	0.1374* [0.0792]	-0.0147 [0.0548]	0.002 [0.0651]	-0.0118 [0.0587]	-0.0302 [0.0598]	-0.021 [0.0599]
Science	0.1987 [0.2250]	0.2961 [0.2569]	0.3891 [0.2781]	-0.0654 [0.2742]	0.174 [0.2730]	-0.0501 [0.2649]	0.3504 [0.2932]	-0.2006 [0.2969]	-0.361 [0.3244]	0.005 [0.3550]	-0.0393 [0.2459]	-0.1936 [0.2923]	0.103 [0.2635]	-0.1795 [0.2678]	0.1182 [0.2688]
Math_Test	-0.0141 [0.2762]	-0.3297 [0.3171]	0.1362 [0.3434]	0.0417 [0.3370]	0.0972 [0.3353]	0.6605** [0.3251]	0.467 [0.3619]	0.7800** [0.3665]	0.7419* [0.3987]	0.6472 [0.4359]	0.2389 [0.3015]	0.294 [0.3608]	0.1055 [0.3254]	0.409 [0.3291]	0.1915 [0.3300]
Correct	0.0985*** [0.0122]	0.0937*** [0.0194]	0.0885*** [0.0211]	0.1139*** [0.0208]	0.1031*** [0.0202]	0.0653*** [0.0144]	0.0759*** [0.0222]	0.0614*** [0.0225]	0.0839*** [0.0246]	0.0385 [0.0263]	0.0583*** [0.0109]	0.0405* [0.0221]	0.0405** [0.0200]	0.0693*** [0.0203]	0.0704*** [0.0199]
Obs.	608	152	152	152	152	608	152	152	152	152	608	152	152	152	152
No. of subject	152					152					152				

Notes: * denotes significance at the 10% level, ** denotes significance at the 5% 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 *Science_Degree* takes value 1 if the degree of study is Architecture, Biology, Business, Computer Science, Economics, Engineering or Human Resources and 0 if the degree is 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”.

Table 6: Gender Differences on the Treatment Effect on Affective Response: 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	4.8937***	4.8301***	5.0566***	4.2652***	4.9812***	4.1817***	3.8935***	3.8628***	3.1942***	3.3278***	5.4350***	5.5046***	5.7919***	5.1789***	5.1576***
	[0.2874]	[0.3456]	[0.4252]	[0.4673]	[0.4782]	[0.3424]	[0.4092]	[0.4537]	[0.5617]	[0.6212]	[0.3002]	[0.3977]	[0.4040]	[0.4649]	[0.4734]
Treatment	-0.2658	-0.4237	-0.2731	-0.206	-0.155	0.1103	-0.2444	-0.1386	0.3573	0.4574	-0.0135	0.0701	-0.1103	-0.0951	0.271
	[0.3381]	[0.4050]	[0.4297]	[0.4169]	[0.4136]	[0.4044]	[0.4795]	[0.4585]	[0.5012]	[0.5372]	[0.3717]	[0.4660]	[0.4083]	[0.4148]	[0.4093]
Girl	0.1607	0.3432	0.0584	0.3888	-0.1369	0.1685	-0.1708	0.2693	0.276	0.2906	-0.0531	0.1578	-0.1179	-0.0733	-0.0548
	[0.3041]	[0.3531]	[0.3845]	[0.3725]	[0.3696]	[0.3638]	[0.4180]	[0.4102]	[0.4478]	[0.4801]	[0.3358]	[0.4062]	[0.3653]	[0.3706]	[0.3659]
GirlxTreatment	-0.0045	0.374	-0.0841	-0.3238	0.0089	-0.0535	0.3685	0.161	-0.5427	-0.2046	-0.2552	-0.1239	-0.2104	-0.2913	-0.5803
	[0.4237]	[0.4957]	[0.5359]	[0.5206]	[0.5182]	[0.5069]	[0.5868]	[0.5717]	[0.6259]	[0.6732]	[0.4672]	[0.5703]	[0.5091]	[0.5180]	[0.5129]
Period	-0.035					-0.2614***					-0.0297				
	[0.0430]					[0.0505]					[0.0352]				
Correct	0.1004***	0.0975***	0.0907***	0.1143***	0.1012***	0.0698***	0.0808***	0.0650***	0.0844***	0.0444*	0.0570***	0.0367	0.0385*	0.0662***	0.0636***
	[0.0121]	[0.0195]	[0.0211]	[0.0206]	[0.0202]	[0.0143]	[0.0231]	[0.0225]	[0.0247]	[0.0263]	[0.0107]	[0.0224]	[0.0200]	[0.0205]	[0.0200]
Observations	640	160	160	160	160	640	160	160	160	160	640	160	160	160	160
No. of subject	160					160					160				

Notes: * denotes significance at the 10% level, ** denotes significance at the 5% 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.

Table 7: Feedback Content on the Treatment Effect on Affective Response: Happiness, Arousal, Dominance

	HAPPINESS				AROUSAL				DOMINANCE			
	ALL	PERIODS 2-4	PERIODS 3-4	PERIOD 4	ALL	PERIODS 2-4	PERIODS 3-4	PERIOD 4	ALL	PERIODS 2-4	PERIODS 3-4	PERIOD 4
Constant	6.2565*** [0.1798]	6.1616*** [0.2653]	5.6948*** [0.4335]	6.4800*** [0.3200]	5.1359*** [0.2099]	5.3832*** [0.2983]	6.3054*** [0.5031]	4.4800*** [0.4020]	6.1323*** [0.1789]	6.1673*** [0.2194]	6.2326*** [0.3276]	6.2400*** [0.3024]
Treatment	-0.2831 [0.2323]	-0.5393* [0.2932]	-0.6411** [0.3269]	-0.3689 [0.4441]	0.1136 [0.2726]	0.2949 [0.3394]	0.191 [0.3958]	-0.2207 [0.5578]	-0.1941 [0.2411]	-0.3627 [0.2680]	-0.6021** [0.2966]	-0.8326** [0.4197]
Positive	0.4070** [0.1629]	0.1812 [0.2062]	0.4696* [0.2501]	0.6473* [0.3860]	0.3823** [0.1869]	0.5942*** [0.2274]	0.4033 [0.2948]	-0.0982 [0.4848]	0.1756 [0.1304]	0.1475 [0.1520]	0.0709 [0.1962]	0.2327 [0.3647]
Positive x Treatment	0.4704** [0.2262]	0.7154** [0.3033]	0.8946** [0.3661]	0.619 [0.5405]	0.2293 [0.2596]	-0.0808 [0.3368]	0.1605 [0.4341]	0.9144 [0.6788]	0.3334* [0.1815]	0.4712** [0.2288]	0.8861*** [0.2943]	1.3221** [0.5108]
Period	0.0805** [0.0407]	0.1595*** [0.0580]	0.2375** [0.1092]	-	0.1888*** [0.0465]	-0.3100*** [0.0625]	-0.5300*** [0.1249]	-	0.0392 [0.0315]	0.0416 [0.0398]	0.0364 [0.0763]	-
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: * denotes significance at the 10% level, ** denotes significance at the 5% 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* becomes positive feedback for the treated subjects while this information is never revealed to the control subjects.

Table 8: Treatment Effect on Performance under Flat-rate Incentives

	Submitted					Correct				
	ALL	PERIOD 1	PERIOD 2	PERIOD 3	PERIOD 4	ALL	PERIOD 1	PERIOD 2	PERIOD 3	PERIOD 4
Constant	14.7115***	16.4231***	19.2949***	21.4487***	22.6667***	12.7372***	14.1538***	17.0128***	19.3718***	20.0641***
	[0.7095]	[0.6393]	[0.7263]	[0.7333]	[0.7464]	[0.7395]	[0.6714]	[0.7541]	[0.7791]	[0.7661]
Treatment	-0.0577	-0.5897	0.6538	0.0128	-0.3077	-0.3333	-0.5513	0.4615	-0.8333	-0.4103
	[0.9556]	[0.9042]	[1.0271]	[1.0370]	[1.0555]	[0.9888]	[0.9495]	[1.0664]	[1.1018]	[1.0835]
Period	2.0987***					1.9654***				
	[0.0865]					[0.0963]				
Observations	624	156	156	156	156	624	156	156	156	156
No. of subject	156					156				

Notes: * denotes significance at the 10% level, ** denotes significance at the 5% and *** denotes significance at the 1% level.

Table 9: Gender Differences in the Treatment Effect on Performance under Flat-Rate Incentives										
	Number of Submitted Summations					Number of Correct Summations				
	ALL	PERIOD 1	PERIOD 2	PERIOD 3	PERIOD 4	ALL	PERIOD 1	PERIOD 2	PERIOD 3	PERIOD 4
Constant	14.5865*** [1.0358]	16.4848*** [0.9696]	19.2727*** [1.0910]	20.8485*** [1.0899]	22.7273*** [1.1251]	12.5259*** [1.0799]	14.4242*** [1.0222]	16.9394*** [1.1378]	18.5152*** [1.1671]	19.8788*** [1.1557]
Treatment	2.6263* [1.4555]	1.2893 [1.3932]	3.3402** [1.5676]	3.6676** [1.5660]	2.2082 [1.6166]	2.3348 [1.5126]	0.9951 [1.4687]	3.0929* [1.6348]	2.8719* [1.6769]	2.3793 [1.6605]
Girl	0.2167 [1.3336]	-0.1071 [1.2766]	0.0384 [1.4364]	1.0404 [1.4349]	-0.1051 [1.4813]	0.3662 [1.3860]	-0.4687 [1.3457]	0.1273 [1.4980]	1.4848 [1.5365]	0.3212 [1.5215]
GirlxTreatment	-4.4636** [1.8951]	-3.1139* [1.8140]	-4.4598** [2.0411]	-6.1097*** [2.0389]	-4.1709** [2.1049]	-4.4435** [1.9695]	-2.5464 [1.9123]	-4.3723** [2.1286]	-6.2124*** [2.1834]	-4.6431** [2.1621]
Period	2.0987*** [0.0865]					1.9654*** [0.0963]				
Observations	624	156	156	156	156	624	156	156	156	156
No. of subject	156					156				

Notes: * denotes significance at the 10% level, ** denotes significance at the 5% and *** denotes significance at the 1% level.

Table 10: Feedback Content on the Treatment Effect on Affective Response: Happiness, Arousal, Dominance

	HAPPINESS				AROUSAL				DOMINANCE			
	ALL	PERIODS 2-4	PERIODS 3-4	PERIOD 4	ALL	PERIODS 2-4	PERIODS 3-4	PERIOD 4	ALL	PERIODS 2-4	PERIODS 3-4	PERIOD 4
Constant	6.5739*** [0.1950]	6.7501*** [0.2575]	6.6286*** [0.4193]	6.3714*** [0.3005]	4.8265*** [0.2131]	5.0830*** [0.2597]	5.2399*** [0.4171]	4.4571*** [0.2942]	6.3863*** [0.1867]	6.5273*** [0.2214]	5.9565*** [0.3251]	6.2857*** [0.2631]
Treatment	-0.3173 [0.2862]	-0.3173 [0.2862]	-0.339 [0.3237]	-0.0381 [0.4068]	-0.2208 [0.2798]	-0.1612 [0.2988]	-0.091 [0.3320]	-0.0762 [0.3983]	-0.2984 [0.2539]	-0.5739** [0.2666]	-0.5731** [0.2855]	-0.4524 [0.3562]
Positive	0.7021*** [0.1672]	-0.304 [0.2151]	-0.0933 [0.2597]	0.6983* [0.4047]	0.1778 [0.1875]	0.1265 [0.2109]	0.2956 [0.2605]	0.5196 [0.3962]	0.3164** [0.1300]	-0.4067** [0.1688]	-0.1897 [0.2052]	0.1794 [0.3543]
Positive x Treatment	-0.1065 [0.2300]	0.3916 [0.2981]	0.1708 [0.3685]	-0.0317 [0.5716]	-0.0746 [0.2579]	-0.2804 [0.2916]	-0.4235 [0.3697]	-0.4839 [0.5596]	0.0272 [0.1786]	0.5504** [0.2328]	0.3752 [0.2914]	0.5984 [0.5005]
Period	-0.0985*** [0.0374]	0.0407 [0.0544]	0.0619 [0.1012]		-0.0441 [0.0423]	-0.1024** [0.0516]	-0.1690* [0.0993]		-0.0615** [0.0282]	0.0203 [0.0397]	0.1570** [0.0735]	
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: * denotes significance at the 10% level, ** denotes significance at the 5% and *** denotes significance at the 1% level.

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 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 RECIEVE ANY PAYMENT.** Thank you.

This experiment is about individual decisions. Pompeu Fabra University has provided funds to carry it out. You will receive 3euros 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 a "Experiment Code" to guarantee that no participant can identify another one by his/her decisions nor 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 3euros plus whatever you earn in the four periods of the experiment.

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

For example:

$$\begin{array}{r} 12 \\ 59 \\ 40 \\ 25 \\ \hline \end{array}$$

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 5 minutes of the period.

You will be paid for each correct solution, 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 3.15euros (3 euros as a show-up fee plus 0.15euros for the correct solution).

Thus, if you solve 25 summations correctly in the four periods, you will earn a total of 6.75euros (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 19.5euros (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.
- In total disagreement
 - In disagreement
 - Neither in disagreement nor agreement
 - In agreement
 - 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.
- In total disagreement
 - In disagreement
 - Neither in disagreement nor agreement
 - In agreement
 - 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.**
- In total disagreement**
 - In disagreement**
 - Neither in disagreement nor agreement**
 - In agreement**
 - In total agreement**
7. Did you take the Math exam during *Selectividad*?
- Yes
 - No
8. What grade did you obtain in your Math exam during *Selectividad*?