The Role of Goals and Feedback in Incentivizing Performance

Zafer Akın[a],* and Emin Karagözoğlu[b]

[a]*Department of Economics, İpek University, Ankara, Turkey*
[b]*Department of Economics, Bilkent University, Ankara, Turkey*

In this paper, we experimentally investigate how goal setting and feedback policies affect work performance. In particular, we study the effects of (i) absolute performance feedback, (ii) self-specified goals, and (iii) exogenous goals and relative performance feedback. Our results show that the average performance of the subjects who are provided self-performance feedback is 11% lower than the ones who obtain no feedback. Moreover, setting a non-binding personal goal does not affect performance. Finally, assigning an exogenous goal and providing relative performance feedback decreases performance by 8%. We discuss the insights our findings offer for the optimal design of goal setting and feedback mechanisms. Copyright © 2015 John Wiley & Sons, Ltd.

1. INTRODUCTION

Incentives are at the core of our lives. Especially in organizations, vast resources are spent to design incentives in order to align the objectives of the organization and the employees. To achieve this, in addition to frequent use of monetary incentives (e.g., bonuses), there are other commonly used methods such as specific goal setting, deadlines, social/peer pressure, and feedback mechanisms. Given the widespread use of these methods in practice, it is crucial to understand how they interact with each other and how people actually react to them. [1]

This experimental study examines the interactive effects of different goal setting and feedback mechanisms on work performance. Specifically, we investigate whether and how self-performance feedback only and self-specified goals combined with self-performance feedback affect work performance.[2] We further study how incentivized assigned-goals interacting with relative performance feedback influence performance.

*Correspondence to: Department of Economics, İpek University, Turan Güneş Bulvarı, 648. Cadde, Oran, 06550, Ankara, Turkey. E-mail: zakin@ipek.edu.tr

Feedback that provides information about employees' performance is a commonly used information sharing method in work environments that allows people to make more informed decisions. Feedback and its effectiveness have been studied mostly in organizational behavior and psychology literature for a long time. However, there is not much of a consensus about how different feedback mechanisms influence employee performance (Balcazar *et al.*, 1985; Kluger and DeNisi, 1996; Locke *et al.*, 1990 and Van-Dijk and Kluger, 2004). There are several features that may jointly determine the effectiveness of feedback such as feedback cues, task properties, and situational and methodological variables. Different combinations of these factors may result in different feedback effects. In this paper, using a lab experiment, we first investigate whether and how merely providing continuous self-performance feedback affects work performance in a context within which there is piece-rate payment for performance and no goals are set for participants.

The use of feedback only mechanisms leads to non-uniform results on performance, which was actually argued by Becker (1978) and Ilgen *et al.* (1979): a

goal without feedback is useless and feedback that does not match an existing goal is of little use.[3] However, this does not automatically imply that every goal with feedback is beneficial. Erez (1977) emphasizes that if feedback is combined with motivational effects of setting specific goals that are challenging but attainable, then it leads to beneficial effects on performance. Falk and Knell (2007) theoretically investigate how people choose goals optimally and show that goals should be high enough to feel the challenge and low enough to prevent feelings of failure. Thus, it can be conjectured that the specific choice of goals will determine whether specified goals supported by feedback produce a positive effect on performance. In this paper, we secondly address the relationship between non-binding (not backed up by any monetary incentives) self-specified goals[4] combined with self-performance feedback and work performance. This comparison allows us to understand the effect of self-chosen, non-binding goals as an internal regulatory/motivational tool on work performance.[5]

Finally, in order to further explore the interaction between goals and feedback in determining performance, we study how and to what extent an exogenously given goal made salient by monetary incentives interacts with relative performance feedback. Locke and Latham (1990) suggest that feedback content and perception of (especially exogenously assigned) goal's attainability are closely related. Negative feedback improves performance as long as the goal is perceived as attainable, and positive feedback may not affect (even deteriorate) performance especially when there is no further incentive such as piece rate.

Relative performance feedback or performance appraisals involve sharing information about where an agent stands in his or her work group. Especially in managerial economics and organizational behavior literature, this topic has been studied extensively (for review articles, see Levy & Williams, 2004; Buunk & Gibbons, 2007). One vein of research is pessimistic about its effects: the management literature (e.g., Milkovich & Newman, 1996) in general concludes that advantages of revealing true performance information to the employee are outweighed by its disadvantages.[6] Eriksson, Poulsen, and Villeval (2009), in a lab experiment, find that regardless of the pay scheme used, feedback does not improve performance and information feedback reduces the quality of the low performers' work.[7] On the other hand, there is a stream of both laboratory and field researches that is more optimistic. In a laboratory experiment, Azmat and Iriberri (2012) find a significant and positive effect

of relative performance feedback on performance, independent of the subjects' standings. Hannan et al. (2008) find a positive effect regardless of the content of the feedback under piece-rate incentives.[8] Conflicting implications of the studies in the literature are potentially due to the existence of various design parameters and factors in the decision environment (Kluger and DeNisi, 1996). There are obviously other psychological factors, such as evaluation effect, motivation effect, self-image effect, and sorting effect (Ederer, 2010). In this paper, we finally examine whether the interaction between assigned goals and relative performance feedback has an effect on performance. Participants are assigned an exogenous goal that is incentivized by a two-level piece-rate payment scheme – to make the target more salient – and also provided continuous average group performance. With this information, one can infer not only whether he or she is performing better/worse than the group average but also how far he or she is from the average throughout the experiment. We believe that potentially, group feedback signals whether the target is attainable and works as a moderator variable, influencing the relationship between actual performance and the exogenous target, especially for the ones who lag behind the target.

Our design involves data entry as a real-effort task. It is an attention-focused task, which is ability independent and not enjoyable. It is also repetitive, which mimics real-life jobs that involve repeating the same tasks over time. As mentioned, participants are rewarded according to piece-rate incentives. After completion of each task, feedback information is updated, which differs across treatments.

Our analyses yield some interesting results that are, we believe, driven by different psychological factors. First of all, we find that providing self-performance feedback has a negative effect on performance. This is possibly due to the specific task feature, which requires subjects to stay focused throughout the task completion stage. The piece-rate incentive structure further enforces the need to stay focused. We believe that combining an intense task with frequent self-performance feedback distracted subjects and caused them to focus less on the task compared with the ones who obtained no feedback at all. This allows us to make some inferences about the relationship between feedback frequency and task intensity. Second, in contrast with earlier studies[9] reporting that self-specified goals have positive effects on performance, our results show that self-specified goals do not affect the performance. We believe this result is mainly driven by the

subjects' selection of goals as mere expectation/prediction of how many tasks that can be completed and by not being challenging at all.

Finally, given an exogenously assigned goal, which is supported by a two-level piece-rate payment scheme, we observe that providing relative performance information deteriorates the overall performance compared with the case in which this information is not provided. While this intervention has no effect on the good performers, the overall decline is mainly driven by the worse performances of the ones who could not attain the goal. The combination of relative performance feedback with an exogenously determined goal and its emphasized salience by the two-level piece-rate incentive scheme is the potential reason of this negative effect on the ones who are unable to attain the target.[10] Our result in this case can be connected to the relationship between negative feedback and goal's attainability made salient by our payment scheme.

The rest of the paper is organized as follows. Section 2 introduces the experimental design in detail. Section 3 describes our experimental results. Section 4 concludes with a brief discussion of the implications and limitations of our results.

2. EXPERIMENTAL DESIGN

The experiment was conducted in May 2012 in the experimental lab of TOBB University of Economics and Technology (Ankara, Turkey) with undergraduates from various majors. All sessions were conducted in a computerized environment and lasted approximately 1 h. In total, we had 10 groups and five treatments. We gathered data from 157 subjects. Announcement was made by email, and students signed up a group online. Subjects were identified by their ID number. Once they are registered online, the system automatically recognizes their ID number and email addresses. In addition to a 5 TL show-up fee, the average participant earned 18 TL (total of 23 TL ~$13). After the arrival at the lab, subjects were placed randomly to separate computer stations that prevent seeing others' screens. The experiment started with the instructions, which were read aloud to all subjects (See Appendix for the instructions of TR5). The following stages were a 5-min trial period, the 30-min actual task, and survey, in that order. After the survey, subjects were paid privately, and the experiment ended (Screen shot for TR3 is shown in Figure A1 in the Appendix).

Participants completed a data entry task that is one of the commonly used real-effort tasks in the literature (e.g., Hennig-Schmidt *et al.*, 2010; Gneezy and List, 2006). The task was entering (hypothetical) exam grades of students on the computer screen. There was a set of 20 hypothetical student ID numbers and corresponding grades.[11] On the right of it, the ID numbers were randomly ordered, and the subjects were supposed to find and enter the corresponding grade correctly. If all 20 grades were entered correctly, this was counted as a successfully completed task. After pressing OK, the screen was refreshed; a new set of numbers were displayed, and the information on the right of the screen was updated.

In total, we conducted five treatments. Details about each treatment are summarized in Table 1. Subjects in all treatments always saw the remaining time on their screens. In all treatments except TR1, they were also provided their own performance in terms of their correct answers/entries and the amount earned until that time. Only in TR5 that participants were able to see the average of their group's number of successfully completed tasks in a continuously updated fashion. In all treatments, participants earned money in a linear piece-rate fashion. They earned 1 TL (approximately 55 cents at the time of the experiment) per successfully completed task. Moreover, in TR4 and TR5, after reaching the exogenously assigned target, they earned 50% more per task (1.5 TL). This is sort of a two-level linear piece-rate payoff scheme (because before the target, they earn 1 TL per task, and after the target, they earn 1.5 TL per task). Regarding the target, in TR1 and TR2, subjects did not choose or were not

Table 1. Experimental Design

Treatment	Target	Own feedback	Group feedback	Payoff
TR1	No	No	No	Linear piece rate
TR2	No	Yes	No	Linear piece rate
TR3	Endogenous	Yes	No	Linear piece rate
TR4	Exogenous	Yes	No	Two-level linear piece rate
TR5	Exogenous	Yes	Yes	Two-level linear piece rate

assigned a target number of tasks. On the other hand, in TR3, subjects determined their own target task number that they planned to finish in 30 min task period. Each subject always saw his or her target on the screen during the actual experiment. Subjects in TR3 were asked to specify a target after the 5 min trial period and before the 30 min actual task. In TR4 and TR5, subjects were assigned a target task number (the assigned target is 13). In TR3, the target did not affect the payoff of the agents (they were paid as described whether they reached their target or not).

TR1 can be considered as the base treatment where there is no determined target or feedback at all. By comparing TR1 and TR2, we can investigate the effect of providing feedback about own performance. Comparing TR2 and TR3 allows us to test whether a self-chosen and non-binding target affects the performance compared with a case where no explicit target is set. Because TR4 and TR5 only differ in group feedback dimension, by comparing them, we can identify the effect of revealing the group average (that is potentially interacting with the assigned target) on performance. These are summarized in Table 1. Table 2 summarizes the main comparisons and the investigated effects.

In TR1, only the task number and the remaining time information were shown. In TR2, all information in the screen shot except the target was shown. In TR4, instead of 'Target You Set (Earnings)', the following was shown 'Your Target (Earnings): 13 (TL)'. The exogenous target and corresponding earnings were set as 13 and 13 TL, which is determined on the basis of our analysis of pilot sessions run previously. In TR4 and TR5, because a two-level

linear piece-rate scheme was used, once the subjects reached the target, their earnings were calculated differently afterwards. The screens in TR4 and TR5 were exactly the same except that the subjects in TR5 also saw information about the average number of successfully completed tasks in their group on the bottom right of the screen. In the instructions, all of these details including the type of feedback they would obtain were clearly explained.

The main independent variables we use in our analyses are self-performance feedback, goal setting, and relative performance feedback. We measure and use the completion time of each (both correct and incorrect) task and the completion time of each correct task as dependent variables. Moreover, we used some demographic and personality trait variables as controls in the analysis.

3. RESULTS

In this section, we present our main results. Table 3 shows summary statistics for the average time per completed task (in seconds). As mentioned earlier, we actually look at two different variables, which are completion time per completed task and completion time per successfully completed task. Throughout the paper, in our reported results, dependent variable is the first variable. We repeated all the regressions with the latter. Unless otherwise noted, the results are qualitatively the same. Table 4 summarizes the main comparisons between the treatments in terms of both the average time of completed tasks and average time of successfully completed tasks. Figure 1 shows the distribution of task completion times for each treatment. Averages in each treatment and in the pooled data are also shown in the figure.

In the following subsections, we make pairwise treatment comparisons. In each case, some nonparametric tests and simple regression analyses with aggregate level treatment data are followed by panel data analysis.

Table 2. Main Comparisons

Treatment comparison	Investigated effect
TR1 vs. TR2	Effect of self-feedback
TR2 vs. TR3	Effect of self-imposed target
TR4 vs. TR5	Effect of group feedback

Table 3. Summary Statistics for the Average Time per Completed Tasks by Treatments

Treatment	Obs	Mean	Median	SD	Min.	Max.	25th perc.	75th perc.
1	29	115.03	114	16.7	84	144	106	122
2	30	131.9	130	29.44	95	252	111	142
3	33	133.27	128	26.39	93	188	110	149
4	28	121.71	120	19.12	97	192	107	130
5	36	131.75	129	25.52	78	191	117	146
TOTAL	155	127.2	122	24.37	78	252	110	139

SD, standard deviation.

Table 4. Treatment Comparisons

	TR1–TR2	TR2–TR3	TR4–TR5
Average time of completed tasks	<**	<	<**
Average time of successfully completed tasks	<**	<	<**

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

3.1. The Effect of Self-performance Feedback (TR1 vs. TR2)

Because TR1 is the control treatment in the sense that neither feedback is given nor a target is assigned to the subjects, comparing TR1 with TR2 gives us the effect of giving self-performance feedback. The average task completion times in TR1 and TR2 are 115 and 131.9 s, respectively. Average completion time is less (approximately by 11%) in TR1 than in TR2. This difference is statistically significant (Wilcoxon rank-sum test, $p = 0.014$). Moreover, we also tested the equality of distributions of *average task completion time* using a two-sample Kolmogorov–Smirnov test between TR1 and TR2: the null hypothesis of equality of distributions is rejected ($p = 0.054$; for *average successful task completion time*, it is marginally rejected, $p = 0.10$).

We then run simple OLS regressions by taking TR1 as the base treatment. The results are presented in Table 5. In model 1, we compare the effects of each treatment on the completion time of tasks. The coefficient of the dummy variable for TR2 is positive and significant. This supports the result of our nonparametric analyses that compared with TR1; task completion time is significantly higher in TR2. Model 2 additionally takes into account the performance in the trial period. It turns out that the performance in the trial period is a good predictor of the actual performance and other coefficients are almost quantitatively the same as in model 1 and significant. Model 3 also includes some control variables such as gender, grade point average (GPA), and competitiveness. Our previous results are confirmed in model 3, as well. Signs of most of the control variables are as predicted, but none of them are statistically significant.[12]
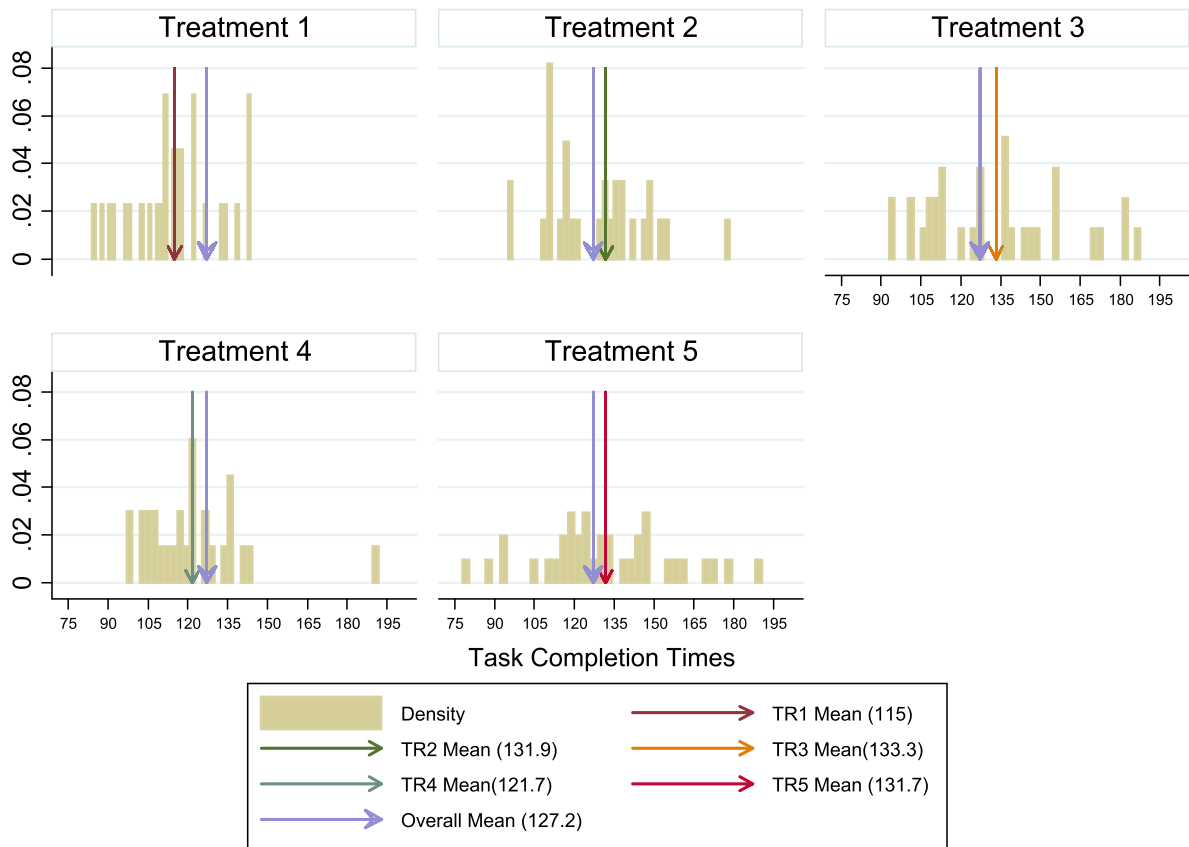

Figure 1. Distribution of Task Completion Times. [Color figure can be viewed at wileyonlinelibrary.com]

Table 5. OLS Estimation – Completion Time

Dependent variable	Completion time		
	Model 1	Model 2	Model 3
TR2	16.86[***] (6.201)	15.92[***] (5.659)	14.95[**] (5.907)
TR3	18.23[***] (5.545)	19.39[***] (5.100)	16.942[***] (5.269)
TR4	4.076 (4.042)	3.367 (4.325)	1.917 (4.630)
TR5	16.71[***] (5.271)	17.52[***] (4.999)	17.65[***] (5.214)
Trial		−12.61[***] (2.262)	−11.32[***] (2.091)
Controls[a]	No	No	Yes
Constant	115.03[***]	130.69[**]	140.37[***]
Number of obs	155	155	155
R^2	0.095	0.273	0.314
P-value of F-test	0.000	0.000	0.000

Notes: Robust standard errors are in parentheses.
[a]Control variables: gender (−4.665, $p = 0.27$), grade point average (0.423, $p = 0.85$), competitive (−1.25, $p = 0.23$), preventive (1.383, $p = 0.19$), and promotion (−1.264, $p = 0.34$).
*, **, and *** Indicate statistical significance at the 0.10, 0.05, and 0.01 level, respectively.

In order to fully utilize the detailed data we obtain, we finally run a panel regression by using all the task completion times of each subject. We run random effects model in order not to lose variables that are fixed (e.g., several individual characteristics) and to capture the treatment effects.[13] Random effects are at the subject level. We pool (unbalanced) data from TR1 and TR2 and estimate the following equation to quantify the effect of giving self-performance feedback (TR2 is a dummy variable whose value is 1 for treatment 2) on the completion time of subject i task j, (Completion Time)$_{ij}$.

$$(Completion\ Time)_{ij} = \beta_0 + \beta_1 TR2 + \beta_1 Trend \qquad (1)$$
$$+ X'_{ij}\delta + \varepsilon_{ij}.$$

The preceding equation is the most inclusive model we run, that is, model 5. Models 4 and 5 include a linear trend. Additional controls (X) include variables

such as ability, errors made, gender, and competitiveness. All the details can be found in Table 6.

In model 1, we include neither linear trend nor any other controls. According to the random effect estimator, having self-performance feedback (i.e., TR2 variable) has a highly significant effect on the completion time, and this effect stays significant in all the models although its magnitude changes a little. Overall, the marginal effect is about 17 s, which corresponds to approximately 11% decline in performance. Model 2 adds the number of successfully completed tasks in the trial period (variable called 'trial') to model 1 as a proxy for the ability of the subjects.[14] While the coefficient of TR2 is almost the same, the corresponding coefficient is negative and statistically significant, which implies that trial period performance is a good predictor of the actual performance.[15] Model 3 controls for the errors that the subjects made

Table 6. Panel Analysis of Treatments 1 and 2

Dependent variable	Completion time				
Model	(1)	(2)	(3)	(4)	(5)
TR2	17.36[***] (5.850)	16.40[***] (5.292)	16.29[***] (5.332)	15.92[***] (5.157)	13.24[***] (4.736)
Trial		−10.78[***] (3.562)	−10.88[***] (3.359)	−10.32[***] (3.123)	−12.26[***] (2.945)
Error			3.094 (9.887)	3.324 (10.17)	3.117 (9.948)
Trend				−0.369 (0.287)	−0.361 (0.274)
Controls[c]	No	No	No	No	Yes
Constant	114.59[***] (3.104)	128.08[***] (6.173)	125.51[***] (13.17)	127.07[***] (14.115)	123.99[***] (25.466)
Observations	858	858	858	858	858
Clusters	59	59	59	59	59
Prob > F	0.003	0.000	0.000	0.000	0.000

Notes: [a]Robust standard errors are in parentheses.
[b]To check whether there is a significant difference across subjects, we run LM test whose null hypothesis is that variances across entities is zero (no panel effect). We repeat this for every regression. In every case, we reject the null hypothesis ($p = 0.000$). Thus, it is appropriate to run a random effect model rather than a simple OLS regression.
[c]Control variables: gender (−14.979, $p = 0.03$), aid (−1.996, $p = 0.71$), competitive (2.82, $p = 0.10$), bored (0.649, $p = 0.58$), preventive (−0.25, $p = 0.81$), and promotion (0.89, $p = 0.75$).
*, **, and *** Indicate statistical significance at the 0.10, 0.05, and 0.01 level, respectively.

in the actual task, but it is not significant. Model 4 further adds a linear trend (warming-up effect) to capture the general evolution of task completion time over the course of the experiment. However, this is not statistically significant either. Finally, in model 5, we add more control variables such as gender and competitiveness.[16] None of the coefficient estimates of these control variables is significant, except gender (male is approximately 15 s faster than females). To sum up, the panel data analysis confirms the previous results obtained from nonparametric analysis. These findings are summarized in our first result:

> Result 1: Under a piece-rate payment scheme, in an intense task with time–pressure, providing feedback about self-performance mitigates performance compared to the case where no feedback is provided.

We believe that potential reasons for this seemingly negative result are the following. Firstly, providing feedback or any other (even potentially useful) information during an intense task may distract attention and divert focus to progress rather than the task itself. In the first treatment, subjects intensively focus on the task because the only available information (other than the task itself) on the screen is the time left. On the other hand, subjects are provided information on their progress in TR2. Because this information is continuously present on the screen during the task, it might have distracted subjects' attention from the task.[17] Secondly, there is potentially an *uncertainty effect*. In TR2, subjects are informed about how many tasks they (successfully) complete and how much they earn at any point during the experiment. This might have created a feeling of progress and a satisfaction associated with it, which may lead to a more relaxed task behavior. On the other hand, when subjects do not receive any other information other than time left, they are under time pressure and they do not know how well they were doing until that time, which may lead them to put more effort to complete as many tasks as they can. Another reason may be the wealth effect, which is likely to be present in experiments where participants can observe their accumulated earnings throughout the experimental session. The presence of a wealth effect would imply that if a participant has a reference earning in his or her mind, his or her performance may fall as he or she observes his or her accumulated earnings approaching to this reference earning. In TR2, our participants can continuously observe how much they earned and how it accumulates as they proceed. Therefore, their performances might

have been affected negatively by the presence of updated progress/earning information.

3.2. The Effect of a Self-chosen Target (TR2 vs. TR3)

Average task completion times in TR2 (no assigned or self-chosen target) and in TR3 (subjects choose their own targets) are 131.9 and 133.27 s, respectively. This difference is not statistically significant (Wilcoxon rank-sum test, $p = 0.77$). The null hypothesis of the equality of distributions between TR2 and TR3 cannot be rejected either ($p = 0.629$).

To further test this *no difference* result, after running OLS regressions (Table 5), for each model, we test whether the coefficients of the dummy variables for TR2 and TR3 are identical. We cannot reject the null hypotheses that these two coefficients are identical, which confirms our previous nonparametric test results (in models 1, 2, and 3, p-values are 0.846, 0.563, and 0.742, respectively).

We then pool the (unbalanced) data from TR2 and TR3 and estimate Equation (1) to quantify the effect of the self-chosen target (TR3 is the dummy variable whose value is 1 for treatment 3) on the completion time of task j of subject i, (*Completion Time*)ij. All the details can be found in Table 7. The coefficient of TR3 is insignificant in all the models.

In model 1, the presence of a self-chosen target has a positive but insignificant effect on the completion time. Model 2 adds the number of successfully completed tasks in trial period to model 1, 'trial' variable, as a proxy for the ability of the subjects. While the coefficient of TR3 changes its sign, it is still not significant, whereas the corresponding coefficient estimate of trial is negative and statistically significant. Model 3 controls for subjects' *errors*, which turn out to be negative, insignificant, and do not change the previous coefficients. Model 4 with the linear trend captures the learning (or warming up) effect. It shows that the coefficient estimate is statistically significant but that this warming up effect is weak. Finally, in model 5 with the inclusion of other control variables, signs and significances of the main coefficients do not change. The coefficients of the control variables are not significant again except for gender with approximately the same effect. To sum up, the panel data analysis supports the previously found results from nonparametric analysis.

In order to understand why there is no significant difference between making subjects setting an endogenous non-binding target and not setting it, we take a

Table 7. Panel Analysis of Treatments 2 and 3

Dependent variable	Completion time				
Model	(1)	(2)	(3)	(4)	(5)
TR3	3.09 (5.850)	−3.48 (6.69)	−3.46 (6.718)	−2.98 (6.369)	−1.36 (4.736)
Trial		−17.73*** (4.683)	−17.69*** (3.359)	−16.76*** (4.303)	−14.75*** (4.130)
Error			−1.60 (7.300)	−1.32 (7.539)	−1.26 (7.495)
Trend				−0.493** (0.252)	−0.413* (0.243)
Controls[a]	No	No	No	No	Yes
Constant	129.02*** (5.472)	156.21*** (10.324)	157.60*** (14.19)	158.57*** (13.674)	176.64*** (25.371)
Observations	847	847	847	847	847
Clusters	61	61	61	61	61
Prob > F	0.667	0.000	0.000	0.000	0.000

Notes: Robust standard errors are in parentheses.
[a]Control variables: gender (-13.56, $p = 0.017$), aid (-8.757, $p = 0.14$), competitive (0.88, $p = 0.53$), bored (2.13, $p = 0.14$), preventive (-2.32, $p = 0.11$), and promotion (-2.43, $p = 0.38$).
*, **, and *** Indicate statistical significance at the 0.10, 0.05, and 0.01 level, respectively.

closer look at the targets subjects set and their actual performances. Firstly, data clearly reveals that subjects used their performance in the trial period when setting their targets. The higher the successfully completed tasks in the trial period, the higher the set target (Spearman rank correlation test: $r = 0.388$ and $p = 0.025$). Secondly, when we compare the number of tasks completed successfully and targets set, we see that they are very close. The mean of the target[18] is even less than the mean of the number of successfully completed tasks (11.78 and 12.43, respectively), but they are not significantly different (Wilcoxon rank-sum test, $p = 0.7342$). Medians do not differ significantly either ($p = 0.608$). Moreover, the hypothesis that the distributions of these two are equal cannot be rejected (Kolmogorov–Smirnov test, $p = 0.762$). Figure 2 plots a kernel density estimate for both of these variables, which indicate that the actually completed tasks and the initially set targets are very

Figure 2. Kernel Distribution for "Target" and "Number of Successfully Completed Tasks". [Color figure can be viewed at wileyonlinelibrary.com]

closely related. These findings are summarized in Result 2:

> Result 2: Under a piece-rate payment, self-performance feedback scheme, and not incentivized, self-chosen targets, the presence of targets does not make any difference in terms of performance with respect to the case where no explicit target is stated.

What we can infer from this result is that subjects seem to have used setting a target as a mere (realistic and somewhat conservative) guess/expectation of how many tasks they can complete. In this sense, targets did not play the role of an actual 'goal', even though subjects see the targets in the screen throughout the actual experiment. Thus, two potential reasons for the *no difference* result are that targets were mere expectations and were not challenging (for most of the subjects) and the fact that targets were not incentivized at all.
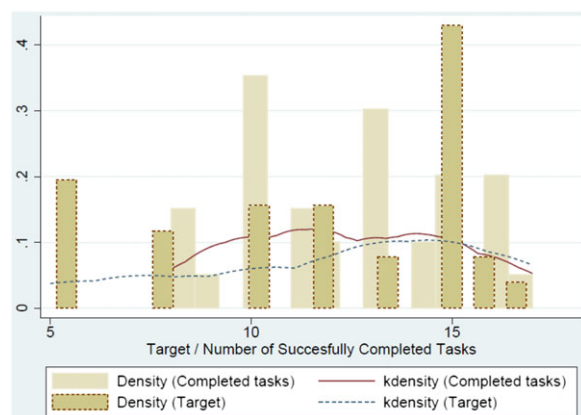
3.3. The Effect of Providing Relative Performance Feedback (TR4 vs. TR5)

Treatments 4 and 5 are different from the previous three treatments in terms of both target and payoff structure. Thus, we can only compare them with each other. In these treatments, we assign, at the beginning of the actual task, a target number of successfully completed tasks for each subject (13, same for everyone). Moreover, subjects are paid at a two-level piece-rate fashion depending on whether they reach this exogenously assigned target or not. Until they reach the target, they earn 1 TL for each successfully completed task, and once they reach the target, they start earning 1.5 TL.

The mean of average completion time (of each subject) in TR4 (121.71) is approximately 8% less than in TR5 (131.75). This difference is statistically significant (Wilcoxon rank-sum test, $p = 0.043$). The null hypothesis of equality of distributions is rejected (Kolmogorov–Smirnov test, $p = 0.062$). These show that revealing information about how others are doing leads to a deterioration of the overall performance. In the OLS regressions (Table 5), we tested whether the coefficients of the dummy variables for TR4 and TR5 are the same for each model. In all models, we reject the null hypothesis of equality of coefficients (In models 1, 2, and 3, p-values are 0.012, 0.004, and 0.0004, respectively). Figure 3 shows the differences in the mean completion times in TR4 and TR5.

In order to explore the data in more detail and see whether the effects mentioned earlier are robust after controlling different variables, we again run a random
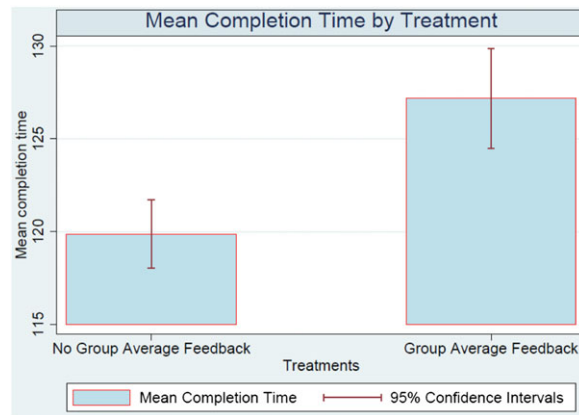

Figure 3. Mean Completion Times in TR4 and TR5 (by Using Panel Data). [Color figure can be viewed at wileyonlinelibrary.com]

effects panel regression by pooling data from treatments 4 and 5. We first estimate, again, Equation (1) to see the effect of relative performance feedback. All the details can be found in Table 8.

In model 1, TR5 has a positive but marginally significant effect on the completion time ($p = 0.076$), and this effect remains positive and significant in all models. Overall, the marginal effect is about 10 s, which corresponds to approximately 8% lower performance in TR5. Model 2 adds *trial* variable. While the coefficient estimate of TR5 is higher (11 s), the coefficient estimate of *trial* is negative and statistically significant. This intuitively shows that as the number of correct answers increases in the trial period, the average completion time decreases. Model 3 controls for the errors subjects made. Its coefficient estimate is positive but not significant, and its inclusion does not change the previous coefficient estimates at all. In model 4 with a linear trend, *warming-up effect* is strong as the coefficient estimate of the trend is negative and highly significant. Adding more control variables affects neither the significance nor the magnitude of the other coefficient estimates. None of the coefficients of these control variables are significant except competitiveness with the expected sign (i.e., negative). To sum up, the results from panel data analyses are parallel with the results from nonparametric analyses. These findings lead to our third result:

Result 3: Under a two-level piece-rate payment and self-performance feedback scheme, providing relative performance information (group average) worsens the overall performance.

A potential reason for this result comes from the theory suggested by Locke and Latham (1990): as long as the given negative feedback does not trigger

Table 8. Panel Analysis of Treatments 4 and 5

Dependent variable	Completion time				
Model	(1)	(2)	(3)	(4)	(5)
TR5	9.74[*] (5.496)	11.15[**] (5.288)	11.12[**] (5.302)	10.35[**] (4.874)	10.63[**] (4.44)
Trial		−10.89[***] (3.320)	−10.92[***] (3.359)	−9.832[***] (3.010)	−7.56[***] (2.579)
Error			2.965 (4.621)	3.901 (4.754)	4.485 (4.773)
				−1.23[***] (0.275)	−1.09[***] (0.270)
Controls[a]	No	No	No	No	Yes
Constant	122.12[***] (3.510)	134.95[***] (5.271)	132.36[***] (7.515)	137.83[***] (7.66)	161.41[***] (20.89)
Observations	904	904	904	904	904
Clusters	64	64	64	64	64
Prob > F	0.076	0.003	0.000	0.000	0.000

Notes: Robust standard errors are in parentheses.
[a]Control variables: gender (−2.32, $p = 0.65$), aid (−6.94, $p = 0.14$), competitive (−3.39, $p = 0.012$), bored (−0.7, $p = 0.60$), preventive (1.56, $p = 0.42$), and promotion (−1.44, $p = 0.31$).
*, **, and *** Indicate statistical significance at the 0.10, 0.05, and 0.01 level, respectively.

the perception of assigned goal's non-attainability, it improves performance. On the other hand, positive feedback is expected at least not to enhance performance especially if it is not further incentivized, such as piece rate.

In order to further investigate the effect of group average feedback, from which subjects can infer their relative standings in their group, we separate subjects in each treatment into two groups: the ones who have numbers of successfully completed tasks above and below the average.[19] If we compare the performance of the people who are above average in TR4 and TR5, we see that the average completion time of subjects in TR4 is less than in TR5 (111.8 vs. 114.2 s). However, the difference is not statistically significant (two-sample Wilcoxon rank-sum test, $p = 0.49$). On the other hand, the average performance of the people who are below average in TR5 is significantly less than the average performance of the ones who are below average in TR4 (125.9 vs. 145.7 s, $p = 0.0063$).[20]

In addition, when we look at the distributions of task completion times in TR4 and TR5, we observe that the dispersion in TR5 is higher than in TR4. We test this by comparing standard deviations of completion times across these treatments and find a significant difference between them (overall $p = 0.0006$ if we ignore one outlier in TR4; for the ones below average $p = 0.013$ and for the ones above average $p = 0.059$).

This implies that revealing relative information makes the performance more scattered and less steady across agents, especially for the ones who are below average. The equality of distributions is also rejected (Kolmogorov–Smirnov test, $p = 0.062$).

In order to capture the differences between participants with performances above and below average more clearly, we run the following panel regression. All the details can be found in Table 9.

$$
\begin{aligned}
(Completion\ Time)_{ij} = {} & \beta_0 + \beta_1 TR5 + \beta_2 Belowave \\
& + \beta_3 Belowave*TR5 \\
& + \beta_4 Trend + X'_{ij}\delta + \varepsilon_{ij}. \quad (2)
\end{aligned}
$$

Belowave is a dummy variable whose value is 1 if the number of successfully completed tasks of an agent (updated continuously when agent completes a task successfully or not) is lower than the average number of successfully completed tasks of the group in the pooled data. Belowave*TR5 is the interaction variable, which is equal to Belowave variable for only TR5. In model 1, in addition to trend, we control for ability and error. The coefficient estimate of the treatment dummy (β_1) is positive and significant ($p = 0.063$). This implies that subjects perform better in TR4, which confirms previous results. The coefficient estimate of Belowave (β_2) is positive and highly significant, which is intuitive. The interaction variable's coefficient estimate tells us whether there is a difference between the below average subjects' performances across treatments. Although this coefficient (β_3) is positive as we found in the previous

Table 9. Effect of Feedback Type and Reaching Target (TR4 and TR5)

Dependent variable	Completion time			
Model	(1)	(2)	(3)	(4)
TR5	7.80[*] (4.197)	8.21[*] (4.323)	12.89[**] (6.362)	17.57[***] (5.730)
Belowave	14.97[***] (4.759)	11.70[***] (4.170)		
Belowave*TR5	6.24 (6.389)	6.80 (5.922)		
Attained			−17.56[***] (5.820)	−12.83[**] (6.181)
Attained*TR5			−12.63 (7.933)	−16.02[***] (7.273)
Trial	−7.75[***] (2.506)	−6.61[***] (2.312)	−6.45[***] (2.090)	−5.82[***] (1.942)
Error	8.93[*] (4.89)	8.48[*] (4.934)	6.279 (4.816)	6.241 (4.786)
Trend	−1.30[***] (0.267)	−1.20[***] (0.264)	−0.916[***] (0.257)	−0.897[***] (0.253)
Controls	No	Yes[a]	No	Yes[b]
Constant	126.41[***] (7.165)	143.99[***] (18.314)	139.89[***] (8.019)	140.30[***] (15.872)
Observations	904	904	904	904
Clusters	64	64	64	64
Prob > F	0.000	0.000	0.000	0.000

Notes: Robust standard errors are in parentheses.
[a]Control variables: gender (−3.24, $p = 0.48$), aid (−4.49, $p = 0.27$), competitive (−2.73, $p = 0.016$), bored (−0.65, $p = 0.60$), preventive (1.3, $p = 0.45$), and promotion (−0.22, $p = 0.86$).
[b]Control variables: gender (−3.16, $p = 0.53$), aid (−0.41, $p = 0.90$), competitive (−2.66, $p = 0.01$), bored (1.12, $p = 0.32$), preventive (0.509, $p = 0.74$), and promotion (0.85, $p = 0.52$).
*, **, and *** Indicate statistical significance at the 0.10, 0.05, and 0.01 level, respectively.

analysis, it is not significant ($p = 0.328$). Thus, we do not find an evidence that supports the previous findings from nonparametric and OLS regression analysis (i.e., the difference between performance of subjects in treatments is driven by the ones who are below average). This result does not change qualitatively when we add the other controls in model 2.

In the goal-setting literature, the goals mainly play a motivational role; and they are generally considered as reference (or anchor) points. Moreover, targets that are considered as unattainable decrease motivation/performance and vice versa (but this effect is potentially mitigated by a piece-rate reward scheme). Because we set an exogenous target for the subjects in TR4 and TR5, we expect the target and relative performance feedback to interact in the following way: while relative performance feedback may motivate the frontrunners about attainability of the target and potentially boost their performance, relative performance feedback may make nonattainability of the target more salient for underdogs, which may deteriorate their performance. To explore this interaction, in model 3 in Table 9, we estimate the following equation in which we investigate whether reaching the exogenous target affects the performance across the treatments.

$$
\begin{aligned}
(Completion\ Time)_{ij} = {} & \beta_0 + \beta_1 TR5 + \beta_2 Attained \\
& + \beta_3 Attained*TR5 \\
& + \beta_4 Trend + X'_{ij}\delta + \varepsilon_{ij}. \quad (3)
\end{aligned}
$$

Before the interpretation of regression results, it is worth noting that 80% of subjects either reached the target and were always above average or could not reach the target and were always below average. That is, only 20% were in the margin of reaching the target (they completed 10, 11, 12, or 13.). Thus, by looking at the effect of reaching the target or not on the performance, we can argue whether information about relative performance feedback plays a mediator role.

Variable *Attained* is a dummy variable taking value 1 if the agent completed 13 or more tasks successfully at the end of the experiment. Within a treatment, a big performance difference between the ones who attained and did not attain the target implies a wide distribution of performances and vice versa. In TR5, there is a large and significant difference (the sum of β_2 and β_3) and when other controls are added, it becomes larger. In TR4, this difference is much smaller when other controls are added (β_2 in model 4, Table 9). This

supports the previous findings about the distributions of the performances across treatments.

Between TR4 and TR5, there is no performance difference between subjects who attained the target (the sum of β_1 and β_3 is positive, but the null hypothesis that this sum is zero cannot be rejected, $p = 0.74$). However, there is a significant difference between subjects who could not attain the target across treatments ($\beta_1 = 17.57$ with $p = 0.002$). In TR5, these subjects perform much worse. Thus, we can say that the overall performance difference between two treatments comes from the performance difference of subjects who are not able to attain the target and the potential reason for this is the treatment variable – relative feedback information – in TR5. Figure 4 shows the mean completion times in TR4 and TR5 that are grouped based on whether the target is reached or not. As clearly seen, there is a significant difference between the groups of subjects who did not attain the target in TR4 and TR5 (Wilcoxon rank-sum test, $p = 0.0001$). Moreover, the performances of the ones who attain the target are not significantly different across these treatments (Wilcoxon rank-sum test, $p = 0.259$).

There are potentially two forces that lead to this asymmetric result. First, in psychology literature, it is argued that when goals are perceived as unreachable, individuals may give up (Locke and Latham, 1990), which leads to a lower performance. We argue that during the experiment, the ones who could not attain the target at the end perceive the (negative) relative performance feedback as a signal that the target is unreachable during the experiment and this affects their performance negatively. Second, if one obtains a negative feedback, it is much more likely that this will be followed by a further negative
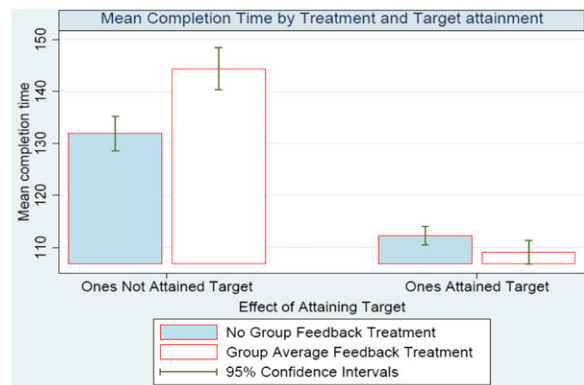

Figure 4. Mean Completion Time by Treatment (4 and 5) and Target Attainment. [Color figure can be viewed at wileyonlinelibrary.com]

feedback (the likelihood of this is[21] 87%). Moreover, the distance from the average for the ones who obtain negative feedback widens over time (Figure 5). This further enforces the perception of the non-attainability of the target.

> Result 4: Under a two-level piece-rate payment and self-performance feedback scheme, the lower performance in TR5 (compared to the ones in TR4) is mainly driven by the low performance of those who could not reach the target.

In order to see whether having a two-level incentive scheme has an effect on the performance, we look at the performance difference between before and after targets among the ones who exceed the target (Table 10). Overall, this does not seem to affect performance. The only difference is in TR5, in that there seems to be a better performance after reaching the target, but this effect is only marginally significant.[22]

Potentially, relative performance feedback may also affect the quality of the subjects' effort. When we look at the error rates across treatments (number of unsuccessful tasks divided by total number of tasks), there is no significant difference between treatments 4 and 5 ($p = 0.197$). When we group the subjects as the ones who attained the target and those who did not, we find some differences (Figure 6). While in treatment 4, between these two groups, there is a slightly significant

Table 10. Effect of Two-level Payoff (TR4 and TR5)

Dependent variable	Completion time	
Model	(1)	(2)
TR5	−0.217 (4.840)	1.40 (4.44)
Abovetarget	−1.87 (3.410)	−0.56 (3.291)
Abovetarget*TR5	−6.81* (4.145)	−6.38 (4.036)
Trial	−3.99* (2.179)	−3.71* (2.215)
Error	12.18** (4.86)	10.13* (5.199)
Trend	−0.67** (0.267)	−0.67*** (0.256)
Controls	No	Yes[a]
Constant	111.76*** (6.322)	141.77*** (14.77)
Observations	493	493
Clusters	31	31
Prob > F	0.000	0.000

Notes: Robust standard errors are in parentheses.
[a]Control variables: gender (−6.09, $p = 0.24$), aid (−2.38, $p = 0.63$), competitive (−2.71, $p = 0.056$), bored (−0.57, $p = 0.72$), preventive (−1.8, $p = 0.19$), and promotion (−0.02, $p = 0.98$).
*, **, and *** Indicate statistical significance at the 0.10, 0.05, and 0.01 level, respectively.

difference ($p = 0.078$), in treatment 5, this difference is high and significant ($p = 0.0003$). Across treatments, while the error rates of the ones who did not attain the target do not differ ($p = 0.651$), for the ones who attained the target, there is a significant difference in that in treatment 5, the error rate is significantly less[23] ($p = 0.024$). What we can infer from these observations is that result 4, mentioning the low performance of the ones who did not attain the target, is not due to their high error rates, but it is due to their slowness in completing the tasks.[24]
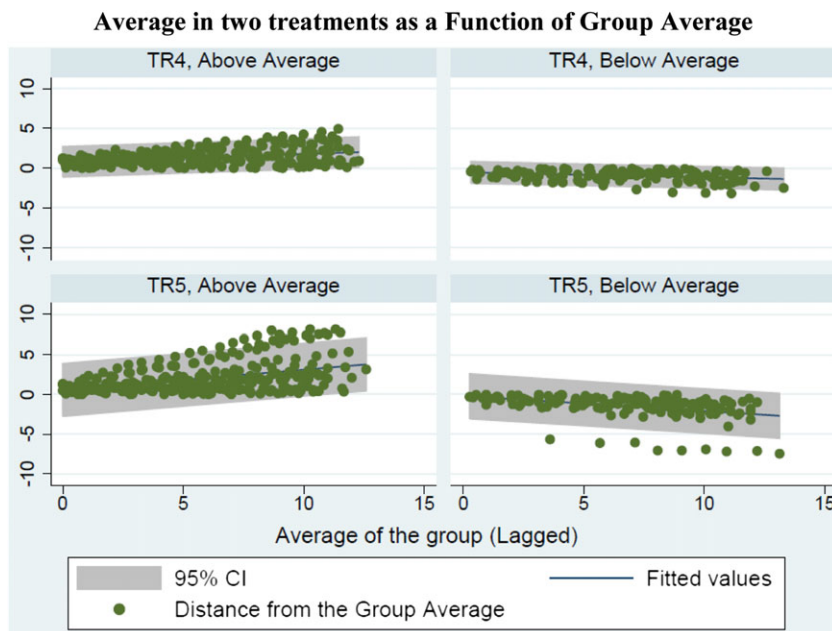

Average in two treatments as a Function of Group Average

Figure 5. Distance from Group Average of the Ones who are Above and Below Group Average in Two Treatments as a Function of Group Average. [Color figure can be viewed at wileyonlinelibrary.com]
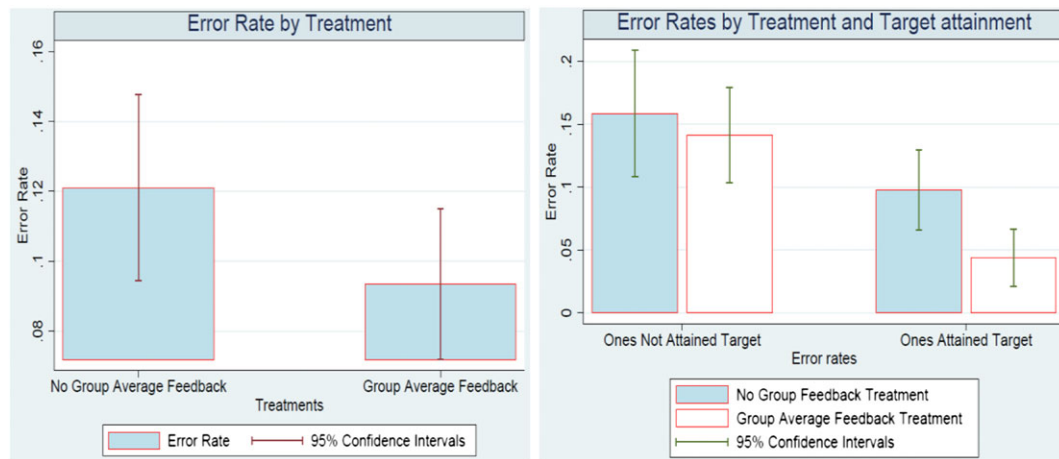
Figure 6. Error Rates. [Color figure can be viewed at wileyonlinelibrary.com]

4. CONCLUSION

In this paper, with five treatments, we experimentally investigate the effectiveness of self-performance feedback, self-chosen targets, and relative performance feedback combined with an exogenous target in increasing work performance in a task that requires steady and careful effort under time pressure. These are very important questions for scholars, managers, policymakers, and practitioners because the methods we investigate here are frequently used in organizations to boost work performance. For successful implementation of these methods, understanding the interactions between them and the environment (e.g., the presence/absence of other methods, task type, and time pressure) is very important. For the firms that are able to provide feedback about performance correctly and easily, these methods are especially important because they allow firms to influence performance in an almost cost-free manner. Our findings show that some feedback methods do not necessarily increase performance as opposed to the common belief. In order for a feedback mechanism to be effective, a thorough perspective should be employed and all the details that may render feedback effective should be carefully considered. Our results add to this understanding by showing that feedback frequency, the absence or presence of extrinsic motivators, and exogenously setting targets play crucial moderator roles in determining the effectiveness of feedback.

In our analyses, we investigate the effects of the aforementioned methods using various non-parametric statistical tests, OLS regressions, and panel regressions, which allow us to see the effects at both individual and aggregate levels and in both a static and a dynamic fashion. First, we observe that providing self-performance feedback does not increase performance and even decreases it (Result 1). This is at odds with the implicit assumption that performance feedback has a positive effect on work performance (for an extensive review of the effects of feedback interventions, see Kluger and DeNisi, 1996, which shows that about 40% of all feedback intervention studies result in negative effect). We think that the presence of a task that requires a steady effort and attention under a time pressure in our experiment offers a reasonable explanation for the negative result. In our experiment, what participants should do to obtain higher earnings is simply to complete as many tasks as possible (correctly) given the time limit. In the presence of such a task, any piece of information that distracts participants' attention from the task likely leads to a deterioration in the task performance: once the feedback information is provided, participants are possibly tempted to look at it, which may reduce their pace, focus, and attention.

Another factor that enforces the previous argument is the screen format that we use. Subjects can always see feedback information on the screen. An alternative design can allow only seeing feedback on the screen at the end of each task for a short period of time, which may potentially mitigate our result if employed. Many real-life tasks – especially in organizations, factories, etc. – are generally completed one after another with an intense focus and under a time pressure. In that sense, we think that the task we have represents a large class of real-life jobs that involve repeating the same tasks over time. To sum up, this result contributes to the literature studying feedback mechanisms by showing that feedback frequency is one of

the important determinants of the effect of the performance feedback.

Second, we observe that the presence of self-imposed targets does not improve task performance (Result 2). This is in contrast with studies that find a positive effect (e.g., Podsakoff and Farh, 1989; Goerg and Kube, 2012). We believe that one major reason for this result is the lack of extrinsic incentives tied to self-imposed targets. In this treatment, participants' earnings are independent of the relationship between their actual performance and self-imposed targets. Hence, participants – when asked to set a target – might have declared just a mere prediction for their performance instead of a *real* target to which they would feel attached. We are aware that even in the lack of extrinsic/monetary incentives, setting a target may influence individual behavior because of intrinsic motivations agents have (Gómez-Miñambres, 2012; Locke *et al.*, 1981; McCalley and Midden, 2002; Wright and Kacmar, 1994). For instance, self-image or social image concerns would be some reasons for such an influence. However, we can infer from our analyses that, in addition to the lack of extrinsic incentives, participants in our experiment possibly lacked strong intrinsic motivations that would positively influence their goal setting and performance as well. This is understandable because (i) it is a very anonymous environment, (ii) there is no interaction (e.g., competition and tournament) with other participants, and (iii) there is no clear reason for subjects to feel any emotional attachment to the task. In summary, this result contributes to the literature that investigates the effect of self-imposed targets on performance, by showing that without necessary tangible or intangible, intrinsic or extrinsic incentives, targets may not be chosen optimally and thus, may not produce the positive effect on performance they are expected to produce.

Third, we observe that providing relative performance feedback combined with an exogenous target (e.g., individual performance and group average together) leads to a deterioration of performance (Result 3). Moreover, this fall in overall performance (at the aggregate level) is mostly due to the fall in the performances of participants who are not able to attain the assigned target (Result 4). This indicates that providing a relative performance feedback may backfire because of the deteriorating morale of low performers. There is already a great debate over the effect of relative performance feedback on individual and/or group performance in organizations. As mentioned earlier, there are studies on both sides, that

is, finding negative and positive effects. In that sense, two results we have here contribute to that strand of literature by presenting yet another instance where relative performance feedback is bad for the overall performance (Bandiera *et al.*, 2013; Ederer and Fehr, 2007; Eriksson *et al.*, 2009; Müller and Schotter, 2010; Pablo and Martínez-Jerez, 2009). In order to clarify the effect of relative performance feedback on performance, a linear piece-rate payoff structure (e.g., nonbinding target) and no assigned target are possible extensions that can be made.

As we mention earlier, the results of our experiment and hence the observed effects of the methods we investigate are crucially influenced by the task characteristics and the presence/absence of other elements such as accompanying extrinsic/material incentives and an exogenous target. We should mention that more experimental studies, some of which are mentioned earlier, where task characteristics and extrinsic incentives are manipulated are needed to be able to conclude that our results are significantly influenced by these factors. Intricate details about the environment, task, time-frame, and so on may gain importance in designing optimal performance enhancement mechanisms.

APPENDIX - INSTRUCTIONS

Thank you for coming. Our experiment involves completing a series of tasks on the computer that require effort and attention. Instructions are simple. If you follow them carefully, you can earn a considerable amount of money. The experiment will last at most 1 hour. Your earnings depend on your task performance. Your earnings will be paid in cash immediately at the end of the experiment. The collected data set in this experiment will be used only for research purposes and the information regarding participants (identity, choice etc.) will be kept completely confidential.

The experiment will start when you log in the experiment website and will continue with a 5-minute practice period designed for understanding the nature of the task. Afterwards, you will be given a target to complete in the 30-minute actual experiment, and then the actual experiment will start. At the end of the experiment, you will be asked to fill out a short survey. Then, you will be paid your earnings personally, and the experiment will be ended.

Here are the task details: As in the following picture, you will be given a series of numbers (20 seven-digit numbers on the left and one, two or three

digit numbers on the right) in a picture format. On the right of this, you will be given the same set of 20 numbers but randomly ordered and be asked to enter the corresponding one, two or three digit number on the right. The entry of this set of numbers correctly is considered as one task. This task is analogous to the following situation: The seven-digit numbers represent student ID numbers and one, two or three digit numbers represent exam grades of these students. You are asked to enter grades of each student into the computer correctly.

After the log in, there will be a practice period so that you can get used to the task and guess more or less how much time you need to complete each task. This practice period will last 5 minutes. At the end of the 5 minutes, you will see a new screen on which you will be given a target to complete in the 30-minute actual experiment. This target will be the same for each participant and affects your earnings from the experiment (this will be explained in the following paragraphs). On the same screen, there will be a "Start the experiment" button. You will need to press this button to start the experiment. It is important for every participant to follow these steps simultaneously in order for the whole experiment to end in a timely manner.

After pressing the "Start the experiment" button, every participant will be able to see the remaining time on the right side of the screen. There will be a number showing the "Task number: …" which will be at the right side of the screen that shows the number of jobs on which you have worked until now. It will be increased by one as you press the OK button at the bottom regardless of whether the task is completed correctly or not. The followings will be also available on the screen: *the target*, *the number of tasks that you have completed correctly until that time*, *and your total cash earnings until that time*. In addition, you will see the average number of tasks that have been completed in your group until that time (this average number will not include your tasks). At the end of 30 minutes, you will be asked to fill out a short survey. After the survey, your total earnings (in TL) will appear on the screen, which will be paid personally to each participant. The amount written on the screen will be the total amount that you will be paid. It will be determined as follows:

Total Earnings:
You will earn (1 TL for every correctly completed task until you reach the target – target included) + (1.5 TL for every correctly completed task above the target) + (5 TL show up fee)

In other words, you will earn 1 TL for each task that you complete correctly in the real experiment. In addition to this, you will earn an extra 0.5 TL bonus for each correctly completed task above the target.

Figure A1. Screen Shot For TR3. [Color figure can be viewed at wileyonlinelibrary.com]

Example: Let the given target be 13. Your earnings from the experiment will be calculated as follows:

If you correctly complete 12 tasks: 12×1 TL $= 12$ TL
If you correctly complete 13 tasks: 13×1 TL $= 13$ TL
If you correctly complete 14 tasks: 13×1 TL $+ (14 - 13) \times 1.5$ TL $= 14.5$ TL
If you correctly complete 15 tasks: 13×1 TL $+ (15 - 13) \times 1.5$ TL $= 16$ TL
If you correctly complete 16 tasks: 13×1 TL $+ (16 - 13) \times 1.5$ TL $= 17.5$ TL

This earning (plus 5 TL show up fee) will be paid to you. You will not be paid for the jobs that are completed in the practice period. The amount of money that you will get will be given on the screen just after the survey. The show up fee will be included in this amount. Completing as many tasks as you can in 30 minutes will increase your earnings.

Some points about which you should be careful.

- Be careful, because if you enter even a single number incorrectly, that whole task will be counted as incorrect.
- Do not press the F5 key, because all numbers that you have entered until that time would be lost and this would cause you to lose time.
- A hint for fast typing: When you want to enter the first number, click on the related place and enter your number. In order to enter the second number, press on the TAB key instead of clicking with your mouse. When you do this, the cursor will move to the next space automatically. This will prevent you to click on spaces every time, and you will be able to save some time.
- To submit a task, click on the 'OK' button. The 'Enter' key will not work.
- While entering numbers, you should enter carefully. When you click on 'OK' button, you will not be asked 'Are you sure?', and you will automatically move to the next task. If you entered all numbers correctly, it will be accepted as correct. However, if you enter even a single number incorrectly, it will be counted as incorrect and this will cause you to lose time.

Thank you for your participation

NOTES

1. This is a topic that has been attracting a great scholarly interest. See Prendergast, 1999; Baker, Jensen, and Murphy, 1988; Murphy and Cleveland, 1995 and Bandiera, Barankay, and Rasul, 2011 for comprehensive reviews of incentive design in firms.
2. Our study mainly utilizes *Feedback Intervention Theory* by Kluger and DeNisi (1996) that is based on both control theory (Carver and Scheier, 1981) and goal-setting theory (Locke and Latham, 1990). It basically argues that feedback interventions change the locus of attention and therefore affect task performance behavior, which is regulated by comparisons of feedback to goals or standards.
3. Heath et al. (1999) argue that goals that people set become their reference points and their behavior overlap with the predictions of prospect theory (Kahneman and Tversky, 1979). Wright and Kacmar (1994) showed that subjects who specify their own goals were less likely to change them when given the opportunity than the ones who were assigned a goal, which indicates a higher commitment to self-set goals. For a thorough review about the relationship between goal setting and feedback, see Locke et al. (1981).
4. In the literature, non-binding and self-specified goals are commonly mentioned as commitment strategies that mitigate self-control problems (see Hsiaw, 2013 and Koch and Nafziger, 2011 for theoretical exposition and Kaur, Kremer, & Mullainathan 2010 for empirical evidence). In our study, commitment aspect of goal-setting does not play a role since our design involves a very short and intense work period that does not have any intertemporal aspect and no self-control issues arise in tasks with these features.
5. In a recent related paper, Gómez-Miñambres (2012) offers a principal-agent model in which the principal uses (non-binding) goals as a tool to manage agents' intrinsic motivation. He incorporates the concept of personal standards into the model, which summarizes the psychological response of the individual towards the goal. He shows that the agents' production, as well as the goals set, increase with the agents' personal standards and goal setting increases agents' achievement and principal's profits in equilibrium.
6. For theoretical models that conclude that it is better to conceal information, see Lazear 1989; Lizzeri, Meyer and Persico, 2003; Ederer, 2010 and Ertac, 2005.
7. See also Bandiera et al. 2013; Delfgaauw et al. (2013); Hannan, Krishnan, & Newman, 2008
8. Azmat and Iriberri (2010), in an educational natural experiment, find that the provision of relative performance feedback enhanced performance by 5% for the whole distribution. Falk and Ichino (2006) observe a very clear positive peer effect on worker performance under a fixed payment scheme. Under flat-rate incentives, Mas and Moretti (2009) find that there are positive

productivity spillovers from the introduction of highly productive personnel into a shift. Blanesi Vidal and Nossol (2011), by using a firm level data set find that individual disclosure of relative standing increased long term average productivity by about 6% without affecting the quality of production.

9. Goerg and Kube (2012), in a field setting, find that both self-chosen goals and exogenously set goals (depending on their size) improve performance even when not backed up by monetary incentives. Podsakoff and Farh (1989) with a two period similar experimental design find that the ones who are informed that they are below average set a higher goal and perform better relative to the ones who are informed that they are above average. This is different from ours in the sense that they not only give feedback before goal setting but also manipulate actual performance information when they give feedback. McCalley and Midden (2002) in an energy conversation setting found that participants in goal-setting with feedback treatments save significantly more energy than those in no-goal-setting with feedback treatment. Goal setting treatments include both a self-set goals group and an experimenter-assigned goals group, and they found that energy savings of these groups are not significantly different from each other.

10. Azmat and Iriberri (2012) is probably the closest study to ours, but there are some important differences. We have one long working period, goal-setting, two-level piece-rate incentive scheme and continuous group feedback. Our approach is also different from Blanes i Vidal and Nossol (2011) in terms of incentive scheme. Bandiera et al. (2013) examine the effect of rank incentives in teams and find that when endogenous team formation is allowed, revealing performance ranking information substantially decreases performance. This difference is mainly due to decline in performances of teams that are below the 40th percentile in the productivity distribution. In our experiment, there is no team structure but the last mentioned result is somewhat related to our result in the sense that the negative effect of relative performance feedback comes from the low performances of people who are not able to attain the goal (more than 80% of whom are always below average). In a field experiment, Delfgaauw et al. (2013) with a tournament scheme find that the ones who are close to winning bonus increase their performances while others do not respond to incentives. However, they find no overall significant effect.

11. Although there is no specific reason to choose this number as 20, we believe that this task induces the subjects to focus more and put more effort into each task, because carelessness costs about 2 min, which is a considerable amount of time in a 30-min experiment. Moreover, we believe that our task also mimics diverse real-life organizational tasks that are completed consecutively under time pressure with an intense focus.

12. The number of observations is not large on aggregate level. Thus, we prefer referring to the results from more reliable panel data regressions.

13. Because subjects are randomly assigned to only one treatment and we do not have subjects participating in different treatments, running the fixed effects model

does not give us the treatment effect. Random effects models are vulnerable to omitted variable bias, but we think we included potentially effective characteristics in the models.

14. Although we carefully choose a task that is relatively ability independent, some people are good at numbers, typing, and so on. The trial variable is added to capture this effect.

15. There is not much variation in the 'trial' variable across subjects. It changes from 0 to 3 with a mean of 1.265 and a standard deviation of 0.87 in treatments 1 and 2. In the whole data, it is virtually the same.

16. Control variables we use in the analysis are gender, GPA, aid, competitive, bored, preventive and promotion. Gender is a dummy variable, 1 for male. GPA is the cumulative grade of subjects out of 4.00. Aid is a dummy variable, 1 for students with scholarship. Competitive is the answer to the question 'How competitive are you?' on a 1–7 scale, 7 being very competitive. *Bored* is rating of 'I am a person who gets bored quickly' on a 1–7 scale, 7 being 'I completely agree.' *Preventive* is rating of 'I was worried that I might not be able to avoid poor outcomes' on a 1–7 scale, 7 being completely agree. *Promotion* is rating of 'I envisioned that I would accomplish desired outcomes' on a 1–7 scale, 7 being 'I completely agree.'. Kluger and DeNisi (1996) identify personality as a moderator of feedback. They conjecture that different personalities may respond to various goal interventions differently. Previous studies use most of the control variables we use. Gender is potentially an important determinant of behavior in this kind of task, in which, although real competition does not exist in all of the sessions, there is a sense of race that may trigger different reactions by different gender. Although the task is relatively ability independent, some users may be more comfortable with numbers and computers. The students with scholarships are in general academically more successful and they may be more able in these kinds of tasks. Aid and GPA, which are strongly correlated, are used due to this possible effect. Competitive measure is again added to capture the possible effect of this characteristic on performance, especially when there is group feedback. Bored variable may be important due to the nature of the task. Getting bored easily may lead to under/overestimation of the results. In previous studies, prevention and promotion focus that are assumed to be the primary motives underlying behavior in self-discrepancy theory (Higgins, 1987) are used both as control variables and as independent variables that potentially affect behavior.

17. In psychology literature, there is substantial evidence that frequent feedback may have negative performance implications (see Kluger and DeNisi, 1996, for an extensive review). Wulf, Schmidt, and Deubel (1993) discuss potential negative effects of frequent feedback in the context of motor learning. Moreover, Salmoni, Schmidt, and Walter (1984) explain negative performance effects of high frequency of feedback on learning transfer of motor tasks and Ilgen, Fisher, and Taylor (1979) discuss how frequent outcome feedback impairs cognitive consistency. Although contexts are different, we believe that same reasoning can be made in our case

such as that introducing frequent feedback may undermine performance.

18. We exclude one subject because although we explained clearly in the instructions, he or she asked explicitly whether the targets affect the payoff and then he or she set an unreasonably high target, 50, which is impossible to reach. He or she actually completed only nine tasks correctly.

19. This is a rough grouping because although most of the subjects were in one of the groups, some of them during the actual task completion were sometimes above and sometimes below the average. We include these subjects in the group in which they spent more time. A more precise panel data analysis will follow.

20. Although we do not present them here, we also run OLS regressions to check whether people above average and below average are affected by provision of relative performance. The results confirm significant difference in the latter and no difference in the former feedback (because the number of observations are small on aggregate level, we prefer referring to the results from more reliable panel data regressions on this analysis).

21. This is because the experiment is a one-shot experiment without separate work periods, and feedback is on the aggregate performance from the beginning to the end of this one-shot experiment.

22. When we run the same regressions with only the tasks that are successfully completed, this weak effect vanishes.

23. We repeated the same analysis for the other treatments too. We made pairwise tests between treatments: the null hypothesis that error rates are the same across treatments cannot be rejected (Kruskal–Wallis test, p-values are greater than 0.40 in all comparisons). We also tested the null hypothesis that the average error rate is equal across all treatments. This hypothesis cannot be rejected either (Kruskal–Wallis test, $p = 0.90$).

24. The numbers of successfully completed tasks in TR4 and TR5 of subjects who could not attain the target are 11.73 and 10.72, respectively, and this difference is statistically significant ($p = 0.043$). For the ones who attained the target, these are 15.22 and 16.27, respectively, and this difference is not statistically significant ($p = 0.667$).

REFERENCES

Azmat G, Iriberri N. 2010. The importance of relative performance feedback information: evidence from a natural experiment using high school students. *Journal of Public Economic* **94**(7–8): 435–452.

Azmat G, Iriberri N. 2012. *The Provision of Relative Performance Feedback Information: An Experimental Analysis of Performance and Happiness*, CEP Discussion Papers dp1116, Centre for Economic Performance, LSE.

Baker GP, Jensen MC, Murphy KJ. 1988. Compensation and incentives: practice vs theory. *The Journal of Finance* **43**(3): 593–616.

Balcazar F, Hopkins BL, Suarez Y. 1985. A critical, objective review of performance feedback. *Journal of Organizational Behavior Management* **7**(3/4): 65–89.

Bandiera O, Barankay I, Rasul I. 2011. Field experiments with firms. *Journal of Economic Perspectives* **25**(3): 63–82.

Bandiera O, Barankay I, Rasul I. 2013. Team incentives: evidence from a firm level experiment. *Journal of the European Economic Association* **11**(5): 1079–1114.

Becker LJ. 1978. Joint effect of feedback and goal setting on performance: a field study of residential energy conservation. *Journal of Applied Psychology* **63**(4): 428–433.

Blanesi Vidal J, Nossol M. 2011. Tournaments without prizes: evidence from personnel records. *Management Science* **57**(10): 1721–1736.

Buunk AP, Gibbons FX. 2007. Social comparison: the end of a theory and the emergence of a field. Organizational Behavior and Human Decision Processes 1021, 3–21.

Carver CS, Scheier MF. 1981. *Attention and Self-regulation: A Control-theory Approach to Human Behavior*, Springer: New York.

Delfgaauw J, Dur R, Non A, Verbeke W. 2013. Dynamic incentive effects of relative performance pay: a field experiment. IZA Discussion Papers 7652, Institute for the Study of Labor (IZA).

Ederer F. 2010. Feedback and motivation in dynamic tournaments. *Journal of Economics & Management Strategy* **19**(3): 733–769.

Ederer F, Fehr E. 2007. Deception and incentives: how dishonesty undermines effort provision. *University of Zurich Institute for Empirical Research in Economics Working Paper*, 341.

Erez M. 1977. Feedback: a necessary condition for the goal setting-performance relationship. *Journal of Applied Psychology* **62**: 624–627.

Eriksson T, Poulsen A, Villeval MC. 2009. Feedback and incentives: experimental evidence. *Labour Economics* **166**: 679–688.

Ertac S. 2005. *Social Comparisons and Optimal Information Revelation: Theory and Experiments*, Mimeo: UCLA.

Falk A, Ichino A. 2006. Clean evidence on peer effects. *Journal of Labor Economics* **24**(1): 39–57.

Falk A, Knell M. 2007. Choosing the Joneses: endogenous goals and reference standards. *The Scandinavian Journal of Economics* **106**(3): 417–435.

Gneezy U, List JA. 2006. Putting behavioral economics to work: testing for gift exchange in labor markets using field experiments. *Econometrica* **74**(5): 1365–1384.

Goerg S, Kube S. 2012. *Goals (th) at Work–Goals, Monetary Incentives, and Workers' Performance*. Preprints of the Max Planck Institute for Research on Collective Goods, No. 2012/19.

Gómez-Miñambres J. 2012. Motivation through goal setting. *Journal of Economic Psychology* **33**(6): 1223–1239.

Hannan RL, Krishnan R, Newman AH. 2008. The effects of disseminating relative performance feedback in tournament and individual performance compensation plans. *The Accounting Review* **834**: 893–913.

Heath C, Larrick RP, Wu G. 1999. Goals as reference points. *Cognitive Psychology* **38**: 79–109.

Hennig-Schmidt H, Sadrieh A, Rockenbach B. 2010. In search of workers' real effort reciprocity—a field and a laboratory experiment. *Journal of the European Economic Association* **84**: 817–837.

Higgins ET. 1987. Self-discrepancy: a theory relating self and affect. *Psychological Review* **94**: 319–340.

Hsiaw A. 2013. Goal-setting and self-control. *Journal of Economic Theory* **1482**: 601–626.

Ilgen DR, Fisher CD, Taylor MS. 1979. Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology* **644**: 349.

Kahneman D, Tversky A. 1979. Prospect theory: an analysis of decision under risk. *Econometrica* **47**(2): 263–292.

Kaur S, Kremer M, Mullainathan S. 2010. Self-control and the development of work arrangements. *The American Economic Review* **100**(2): 624–628.

Kluger AN, DeNisi A. 1996. The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin* **1192**: 254–284.

Koch AK, Nafziger J. 2011. Self-regulation through goal setting. *The Scandinavian Journal of Economics* **1131**: 212–227.

Lazear EP. 1989. Pay equality and industrial politics. *Journal of Political Economy* **97**(3): 561–580.

Levy PE, Williams JR. 2004. The social context of performance appraisal: a review and framework for the future. *Journal of Management* **30**: 881–905.

Lizzeri A, Meyer M, Persico N. 2003. *The Incentive Effects of Interim Performance Evaluations*, mimeo: New York University.

Locke EA, Shaw KN, Saari LM, Latham GP. 1981. Goal setting and task performance: 1969–1980. *Psychological Bulletin* **90**(1): 125.

Locke EA, Latham GP, Smith KJ, Wood RE. 1990. *A Theory of Goal Setting & Task Performance*, Prentice Hall: Englewood Cliffs, NJ.

Mas A, Moretti E. 2009. Peers at work. *American Economic Review* **991**: 112–45.

McCalley LT, Midden CJ. 2002. Energy conservation through product-integrated feedback: the roles of goal-setting and social orientation. *Journal of Economic Psychology* **23**(5): 589–603.

Milkovich GT, Newman JM. 1996. *Compensation*, 5th edn. Irwin: Homewood, IL.

Müller W, Schotter A. 2010. Workaholics and dropouts in organizations. *Journal of the European Economic Association* **84**: 717–743.

Murphy KR, Cleveland JN. 1995. *Understanding performance appraisal: social, organizational, and goal-based perspectives*, Sage: Thousand Oaks, CA.

Pablo CA, Martínez-Jerez FA. 2009. Relative performance compensation, contests, and dynamic incentives. *Management Science* 2009 **55**: 1306–1320.

Podsakoff PM, Farh JL. 1989. Effects of feedback sign and credibility on goal setting and task performance. *Organizational Behavior and Human Decision Processes* **441**: 45–67.

Prendergast C. 1999. The provision of incentives in firms. *Journal of Economic Literature* **37**(1): 7–63.

Salmoni AW, Schmidt RA, Walter CB. 1984. Knowledge of results and motor learning: a review and critical appraisal. *Psychological Bulletin* **95**: 355–386.

Van-Dijk D, Kluger AN. 2004. Feedback sign effect on motivation: is it moderated by regulatory focus? *Applied Psychology: An International Review* **53**(1): 113–135.

Wright PM, Kacmar KM. 1994. Goal specificity as a determinant of goal commitment and goal change. *Organizational Behavior and Human Decision Processes* **59**(2): 242–260.

Wulf G, Schmidt RA, Deubel H. 1993. Reduced feedback frequency enhances generalized motor program learning but not parameterization learning. *Journal of Experimental Psychology Learning Memory and Cognition* **19**: 1134–1150.