

IZA DP No. 7953

**The Power of (No) Recognition:
Experimental Evidence from the University Classroom**

Nicky Hoogveld
Nikolay Zubanov

February 2014

The Power of (No) Recognition: Experimental Evidence from the University Classroom

Nicky Hoogveld

University of Bologna

Nikolay Zubanov

*Goethe University Frankfurt
and IZA*

Discussion Paper No. 7953
February 2014

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

The Power of (No) Recognition: Experimental Evidence from the University Classroom

We study the effect of recognition on performance with a field experiment involving first-year undergraduate students at a Dutch university. Our treatment, given unannounced in randomly selected student groups, was to publicly recognize students who scored within the top 30% of their group on the first of the two midterm exams. The overall treatment effect on the second midterm grade is $0.03s$ (s = the grade's standard deviation) for the recipients of recognition, and $0.15s$ for the non-recipients, both statistically insignificant. The effect for the non-recipients increases with class attendance (itself unaffected), and decreases with the distance to the cutoff grade for recognition, reaching a significant $0.44s$ for those exceeding the minimum attendance requirement and staying within the first quartile of the distance to cutoff. We argue that conformance to performance norm is the most likely behavioral mechanism behind our findings.

JEL Classification: C93, M52

Keywords: recognition, performance, experiment

Corresponding author:

Nikolay Zubanov
Goethe University Frankfurt
Faculty of Economics and Business Administration
Grueneburgplatz 1
60323 Frankfurt am Main
Germany
E-mail: zubanov@econ.uni-frankfurt.de

1 Introduction

Public recognition of employees for good performance is a popular human resource management practice, present in 86% of the U.S. firms surveyed in the 2011 *WorldAtWork* study on trends in employee recognition.¹ Remarkably, in at least 40% of cases recognition involves no cash benefits or career advancement, suggesting that many organizations believe that recognition makes people work harder independently of the material benefits it may bring. This belief speaks to theoretical arguments along the lines of reciprocity (Akerlof, 1982), conditional altruism (Levine, 1998), conformity to social norms (Bernheim, 1994) and status concerns (Moldovanu, Sela and Shi, 2007), and is further supported by a growing body of experimental research (Grant and Gino, 2010; Chen, Harper, Konstan and Li, 2010; Kosfeld and Neckermann, 2011; Kube, Marechal and Puppe, 2012; Bradler, Dur, Neckermann and Non, 2013). Our study contributes to this literature by providing new evidence on how public recognition affects university student performance.

Recognizing student achievement is discussed in major pedagogy textbooks and is recommended as one of the core teaching practices. For example, Marzano, Pickering and Pollock's (2001) *Classroom Instruction That Works*, which sold more than a million copies as of late 2011 and is now into second edition, has a separate chapter on "reinforcing effort and providing recognition" with a summary of relevant educational psychology studies.² Yet, the economics literature has so far paid relatively little attention to recognition as compared to other tools for stimulating academic performance, such as financial incentives to students and teachers, class size reduction, or extra academic support. The only study in this literature we are aware of is Levitt, List, Neckermann and Sadoff (2012) who have found from a series of experiments with Chicago high school students that the effect of a symbolic award promised for improving test score (a trophy and a photo on the wall in

¹Source: <http://www.worldatwork.org/waw/adimLink?id=51194>.

²These studies, most of which are also summarized in Henderlong and Lepper (2002) and Hattie and Timperley (2007), were carried out involving school pupils rather than university students who are the focus of our research.

the class) is positive and comparable to that of financial incentives of up to \$20. Though large for junior students, the award's effect declines into insignificance with student age.

In this study we will try a different way of giving recognition, which relies on spontaneous, unannounced praise of recent performance. We are interested in spontaneous recognition because it occurs frequently in everyday social interactions, within or outside academia, and yet its effects on performance are not well-studied. The other reason for our interest owes itself to the pattern of spontaneous recognition's performance effects, namely, a far greater response from non-recipients than recipients of recognition, found in Bradler et al. (2013) and Chen et al. (2010). In this differential effect on performance, spontaneous recognition stands in stark contrast with pre-announced recognition which, when offered on a competitive basis, triggers a higher response from the more able, as found in Kosfeld and Neckermann (2011). Therefore, when managers or teachers want to improve the performance of their currently underachieving workers or students, spontaneous recognition may work better than a pre-announced award.

Yet, as is true with all incentives, recognition's effect depends on the costs of effort to meet the standard and on the importance of the activity being recognized. Intuitively, the response to recognition will be the weaker, the less important the underlying activity is, and the higher the costs of achieving recognition. As another contribution to the wider literature, we formalize these intuitions in a simple model and test it experimentally.

In Harrison and List's (2004) classification, ours is a "natural field experiment", run in an environment that was normal for the participants, and without them knowing they were part of an experiment. Our core sample includes 342 first-year undergraduate students at a Dutch university attending the compulsory microeconomics course. In the beginning of the course, and without our involvement, these students were divided into 15 tutorial groups each taught by an experienced teaching assistant (TA). Our experimental treatment took place between the two midterm exams during the course, each carrying an equal weight in the

final grade. The treatment, administered through and on behalf of the TAs in 8 randomly chosen tutorial groups, was to give public recognition to students whose grade for the first midterm exam was within the top 30% of their group. We instructed the TAs in the control groups not to express any praise or criticism of the first midterm results in their groups.

We find that, compared to their peers in the control groups, the recipients of recognition in the treatment groups do no better. At the same time, the non-recipients who attend more classes than formally required and are not too far off the cutoff grade for recognition significantly improve their performance. Our findings taken together imply that spontaneous recognition may be an effective motivational tool, but its effectiveness depends on the characteristics of the target group that it cannot influence.

2 Theory

The effect of spontaneous recognition on subsequent performance may be explained within four theoretical perspectives – reciprocity, conditional altruism, status concerns, and conformity to the norm – all of which have received some empirical support. Reciprocity predicts that giving a gift will lead to a positive effort response of the recipients of that gift (Akerlof, 1982; Falk and Fischbacher, 2006). Although recognition is not a material gift (assuming as we do here that there are no material consequences of it), its recipients may still regard it as valuable, and hence reciprocate, when it is costly to the giver (Baron and Kreps, 1999, p. 109; Kube et al., 2012). Conditional altruism (Levine, 1998; Dur, 2009), a related theory, implies that recognition may affect employee effort when it is perceived as the signal that the boss cares about them. Because reciprocity and conditional altruism both require costly recognition, they are not applicable in our case where the effort to provide recognition was negligible. Status concerns and conformity are the remaining possibilities.

Under status concerns, recognition increases effort when receiving it leads to a higher status within the social group. For instance, Moldovanu

et al.’s (2007) “contests for status” model predicts that in the presence of recognition every agent will put in effort proportionate to their ability rank within the group, whereas in the absence of recognition everyone’s effort will be zero. Crucial for this mechanism to work is the expectation that recognition will occur in the future; otherwise, there is no reason for effort to change *after* a one-off recognition.

Conformity to the norm – a tendency to align actions to “a single standard of behavior despite heterogeneous preferences” (Bernheim, 1994, p. 841) – affects effort through feedback on relative performance that comes in the form of recognition. The norm can be operationalized as the median performance (Chen et al., 2010) or, as in Bradler et al. (2013), as the coefficient on effort in the utility function (the higher the norm, the higher the optimal effort). Whatever the norm is, the recipients of recognition will learn that they are more likely to have met the norm than they thought before, and will consequently reduce their effort, whereas the non-recipients will find themselves less likely to comply with the norm, and will therefore work harder. The opposite effort responses by the recipients and non-recipients of recognition is a marker of norm conformity being at work, distinguishing it from status concerns which encourage high performance to work progressively harder. It is what we see in our data, and model in the next section.

2.1 A model of conformity to the norm working through recognition

Consider a group of individuals each of whom cares about their relative performance but does not know their position in the group’s output or ability distribution. Instead, each individual i has beliefs about the percentile to which their ability belongs in the group’s distribution, \tilde{a}_i , summarized by a uniform probability distribution on a $[0, 1]$ interval. That is, each individual’s *ex ante* subjective probability of being within the top $x\%$ of the group’s ability distribution is just $x\%$.

Output is the product of ability (a) and effort (e), $a_i \cdot e_i$, and the costs of effort are $\frac{1}{2} \cdot c \cdot e_i^2$. There is a performance norm $\eta \in [0, 1]$, meaning that each individual considers their output to be within the norm if it

belongs to the top $(1 - \eta) \cdot 100\%$ of the group's output distribution. (This is a generalization of Chen et al.'s (2010) definition of the norm.) As in Bernheim (1994), failure to fulfil the norm brings disutility. The individual payoff function is

$$u_i = \begin{cases} a_i \cdot e_i - \frac{1}{2} \cdot c \cdot e_i^2 & \text{if the norm is fulfilled} \\ a_i \cdot e_i - \frac{1}{2} \cdot c \cdot e_i^2 - \lambda & \text{otherwise} \end{cases} \quad (1)$$

where the term λ is the disutility from failing to meet the norm. The structure of the individual payoff function is common knowledge, except for the ability level a_i known only to individual i .

Each individual separately chooses a level of effort to maximize their own expected payoff function given their beliefs about their relative performance and anticipated effort choices of the rest of the group:

$$\mathbb{E}(u_i) = a_i \cdot e_i - \frac{1}{2} \cdot c \cdot e_i^2 - \lambda + \lambda \cdot \pi(e_i, \mathbf{e}_{-i}), \quad (2)$$

where

$$\pi(e_i, \mathbf{e}_{-i}) = \int_0^1 p(e_i, \mathbf{e}_{-i}, \tilde{a}_i) d\tilde{a}_i$$

is the subjective probability of fulfilling the norm as a function of own effort (e_i) and the anticipated effort levels of the rest of the group (vector \mathbf{e}_{-i}). We assume that $\pi(\cdot)$ is positive, increasing and concave in e_i for $e_i > 0$ and zero for $e_i = 0$. A further inspection reveals that $\pi(\cdot)$ is symmetric, meaning that, given e_i , $\pi(e_i, \mathbf{e}_{-i})$ is equal for all individuals in the group, since the costs of effort are equal for all. This symmetry implies that π 's derivative with respect to own effort taken at the optimal effort level, is equal for all individuals.

In the absence of any information about relative performance or ability, the optimal effort e_i^* is determined from the following first-order condition:

$$a_i - c \cdot e_i^* + \lambda \cdot \pi'_0(e_i^*, \mathbf{e}_{-i}^*) = 0, \quad (3)$$

where $\pi_0(e_i^*, \mathbf{e}_{-i}^*)$ is the subjective probability of fulfilling the norm in this situation. Since $\pi'_0(e_i^*, \mathbf{e}_{-i}^*)$ is the same for all i , the ranking of

optimally chosen effort levels is the same as the ranking of ability levels. Hence, each individual's subjective probability of fulfilling the norm is $\pi_0(e_i^*, \mathbf{e}_{-i}^*) = 1 - \eta$.

Suppose now that there has been a round of public recognition given to the individuals with relative performance $\rho \in [0, 1]$ or above, that is, those whose output is within the top $(1 - \rho) \cdot 100\%$ of the group's output distribution. As a result, both the recipients (R) and non-recipients (NR) of recognition change their subjective probability of fulfilling the norm from π_0 to π_1 as shown in Table 1 below.

[Table 1 here.]

Table 1 shows that, irrespective of the position of the cutoff performance level for recognition ρ relative to the performance norm η , recognition leads the individuals to update their beliefs in such a way that the subjective probability of fulfilling the norm given the effort increases for the recipients and decreases for non-recipients. As an illustration, suppose the norm is the median performance, $\eta = 0.5$, as in Chen et al. (2010), and, as in Bradler et al. (2013), recognition is given to the top 37.5%, that is, $\rho = 0.625$. The R will learn that they have certainly fulfilled the norm, so that their updated probability $\pi_1 = 1$. The NR will learn that they are definitely not within the top 37.5% of the output distribution, but they might still be within the top 50%, and thus fulfilling the norm, the probability of which event is $\frac{0.5-0.375}{0.625} = 0.2$. Given the previously chosen set of optimal effort levels $(e_i^*, \mathbf{e}_{-i}^*)$ and the assumptions of $\pi(\cdot)$'s monotonicity, concavity and being zero at $e_i = 0$, the post-recognition changes in probabilities of fulfilling the norm imply

$$\pi'_{1,R}(e_i^*, \mathbf{e}_{-i}^*) < \pi'_{0,R}(e_i^*, \mathbf{e}_{-i}^*) \quad \text{and} \quad \pi'_{1,NR}(e_i^*, \mathbf{e}_{-i}^*) > \pi'_{0,NR}(e_i^*, \mathbf{e}_{-i}^*) \quad (4)$$

Hence, to satisfy the first-order condition (3) with the updated probability of fulfilling the norm, $\pi_1(\cdot)$, post-recognition optimal effort will increase for non-recipients and decrease for recipients.

The magnitude of effort's response to recognition will depend on the difference between the subjective probabilities of fulfilling the norm pre- and post-recognition, $\pi_1 - \pi_0$, which in turn depends on the recognition

cutoff ρ (see Table 1). An extremely tight or lax cutoff (ρ close to 1 or 0) will produce a weaker response because recognition brings little extra information about one’s relative performance. A less extreme ρ is more informative, and increasingly so as it approaches the norm η , as long as η is not too close to 0 or 1. In fact, $\rho = \eta$ should produce the highest effort response, since then everyone’s performance with respect to the norm will be perfectly known.

A further comparative statics on (3) show that, given η and ρ , the response to recognition will be weakened by larger marginal costs of effort (higher c), and strengthened by higher disutility from failing to meet the norm (higher λ). Thus, the effect of recognition on performance through conformity to the norm is heterogeneous in at least two dimensions: the costs of effort and the individual importance of the activity being recognized. Our experimental data can account for these heterogeneities. There may be other differences between individuals as well, most importantly, in the norm η , which may be incorporated in our baseline model. These extensions will result in the subjective probability of fulfilling the norm before recognition, $\pi_0(e_i^*, \mathbf{e}_{-i}^*)$, to vary by individual. However, since the adjustment of effort by the recipients and non-recipients of recognition does not depend on the disposition of η and ρ , the mechanism generating the opposite responses in effort by R and NR will remain in force. Hence, the model’s predictions will be qualitatively the same, although unobserved heterogeneity will add to the noise in the treatment effect’s estimates.

3 Experiment

Our experiment involved first-year undergraduate students attending the compulsory microeconomics course at a large university in The Netherlands in the winter semester of 2012/13. The course was taught in a way typical for a modern public university, which included regularly scheduled lectures given by the course leader to all students at once, tutorials taught by teaching assistants (TAs) to smaller groups, two midterm exams weighing 10% each in the final grade, and the final exam at the end of the course. Our treatment took place between those two midterms

and was given at the level of tutorial group.

The tutorial groups were formed prior to the experiment and without the involvement of any of the authors.³ Of the 29 tutorial groups on the course, we were given permission by the course leader to run our experiment in 15 – all taught by TAs with at least one year of experience. By random allocation, the treatment was given in 8 groups, and the remaining 7 formed the control part of our sample.

Our treatment was administered with the help of ten group TAs, five of them teaching two groups each and the rest teaching one group. Before the start of the experiment, the TAs were invited to an information session where we explained the purpose of our experiment, requested their participation, and asked them not to tell about the experiment to the students. All TAs agreed to both requests. They were informed whether they were in a treatment or a control group after they had finished grading the first midterm. (The TAs who taught two groups had one treatment and one control group, randomly assigned.) The TAs in the treatment groups were asked to recognize in public the students whose grade for the first midterm was within the top 30% of their group. The TAs in the control groups were asked *not* to give any comments, positive or negative, on the first midterm results.

The procedure, which the TA in the treatment groups were asked to adhere to, is as follows. At the beginning of the first tutorial after the first midterm, the TA informed the students that their exam papers were to be handed out and discussed. Thereafter, he or she said the following, in Dutch: *I would like to have your attention for the students to whom I will now hand out their papers, as they did an excellent job. All of them received at least grade [the cutoff grade corresponding to the top 30% in the respective group, supplied by the authors]. Experience tells us that students find the microeconomics midterms very hard. But these students did very well. My compliments!* After distributing those papers, in a random order, the TA continued by saying *I will now hand*

³One of the authors was a teaching assistant in two of the groups, one treatment and one control. Excluding those groups from the study sample does not change the results in any way.

out the rest of the papers, distributing them in random order as well. The tutorial then proceeded as usual, with each exam question being discussed and the exam papers collected in the end. There was no token or record than came along with the TA's message of recognition. Hence, the recipients could not convert the recognition given to them into future career or status benefits.

[Figure 1 here.]

Figure 1 gives a summary of our experimental procedures in chronological order. It is worth noting that the timing of the events is such that the last tutorial, when the students met their TAs for the last time, took place only 1 or 2 days after the second midterm exam. With this timing known to every course participant through official communications, there should be little expectation among students for the second midterm results to become available in time for recognition to recur. Therefore, while it cannot be completely excluded, the extent to which recognition can affect performance through status concerns is limited.

4 Data

The key outcome variable in the main part of our analysis is the midterm exam grade. In the Dutch education system, grades range from 1 (the lowest) to 10 (the highest). Other potentially interesting outcomes – the final exam grade, presence at the tutorials, the probability of turning up for the midterm – are studied as an extension. The data on all these variables come from administrative records, a highly reliable source.

In addition to the above data, we tried to obtain more information from the students that we could link to their performance, most importantly, the hours spent studying on the course, which we hoped to use as an input measure of effort. For this purpose, we ran the same questionnaire before and after the treatment on behalf of the author not involved in the course, promising that the responses will not be communicated to the course leader (which promise we of course kept). As well as the study hours question, it had questions on gender, year of birth, province of origin, and satisfaction with student life, all of which could provide useful information. However, despite all our efforts, the response rate

on the study hours question was less than 30% and, for those who did respond, the correlation with their grade was about zero, suggesting that the answers may not be reliable. Still, we report the treatment’s effect on self-reported study hours as a side result.

Our complete sample consists of 190 students in 8 treatment groups and 152 students in 7 control groups. There are 75% male and 25% female students, most of whom are aged between 18 and 20. This sample resulted from excluding students not present on the (unannounced) treatment day (20% of the original sample) and those missing one or both midterms (a further 7% of the remaining sample). The latter exclusion was necessary to ensure that all our regression specifications were run on the same sample. Focussing on the subset of students present on both midterms could result in biased estimates of the treatment effect if midterm participation were affected by our treatment. We address this possibility as an extension to our main analysis.

5 Method

We measure the effect of our experimental treatment using the difference-in-difference estimator. In its basic version, it calculates the average difference between the post- and pre-treatment outcomes specifically for the treatment and control groups, estimating the treatment effect as the difference between those two differences. The same point estimate, as well as its standard error, can be obtained from regressing the post- to pre-treatment outcome difference Δy_i for each participant i on the treatment dummy that takes the value of 1 if i was in one of the treatment groups, and 0 otherwise:

$$\Delta y_i = \beta_0 + \beta_1 * treatment_i + error_i \quad (5)$$

Since our theoretical model predicts a different effect of recognition on receivers and non-receivers, we will estimate the treatment effect separately for each group (R = receivers, N = non-receivers), by running

$$\Delta y_i^{R,N} = \beta_0^{R,N} + \beta_1^{R,N} * treatment_i + error_i^{R,N} \quad (6)$$

In equation (6), the would-be receivers (non-receivers) in the control groups are identified as falling within (outside) the top 30% of the first midterm grade distribution in their group.

The estimation issues applicable to both equations (5) and (6) are i) the possible bias to the estimate of the treatment effect (β_1) due to regression to the mean, and ii) the cluster structure of the error term, which invalidates the conventional OLS estimates of the treatment effect's standard error. Regression to the mean is likely to occur when selection into treatment is correlated with the pre-treatment value of the outcome, in which case the natural convergence of the outcomes in the control and treatment groups to a common mean will be mistaken for the treatment effect (Stigler, 1997). To see this, note that (5) is a restricted version of the more general equation

$$\Delta y_i = \gamma \cdot \left(y_{i,-1} - \frac{-\beta_0}{\gamma} \right) + \beta_1 * treatment_i + error_i, \quad (7)$$

where $y_{i,-1}$ is the pre-treatment outcome and parameter γ measures how strongly the outcome distributions in the control and treatment groups converge to a common mean $\frac{-\beta_0}{\gamma}$. The OLS estimate of the treatment effect from (5) is then

$$\hat{\beta}_1 = \beta_1 + \gamma \frac{\text{cov}(y_{i,-1}, treatment_i)}{\text{var}(treatment_i)}$$

which is biased unless $\gamma = 0$ (no convergence) or the average pre-treatment grades in the treatment and control groups are equal. As we show in the next section (Tables 2 and 3), neither of these conditions is satisfied in our data. As a solution to remove the bias to the treatment effect's estimate, we run the generalized difference-in-difference equation (7) in addition to the baseline equations (5)-(6).

The second estimation issue – clustered standard errors – is important because individual error terms in equations (5)-(7) may be correlated within tutorial groups via group-specific unobservables (e.g., time of the day, TA quality, etc.). Because our treatment is also given at the group level, such correlation will invalidate inferences based on the stan-

standard OLS procedure which assume the error terms' independence across individuals. In particular, as shown by Moulton (1986), the treatment effect's standard error will be underestimated by a factor (later named the Moulton factor) proportional to the strength of the within-group correlation in errors and to the number of observations in a group.

Angrist and Pischke (2009, chapter 8) discuss five approaches to repairing the regression coefficients' standard errors in the presence of clustering, of which we implement four:

1. Calculate the Moulton factor from the data and use it to correct the unadjusted standard errors. This solution assumes homoscedasticity in the individual component of the error term.
2. Estimate from the data the residual variance-covariance matrix that allows for within-group residual correlation as well as heteroscedasticity. This solution, proposed in Liang and Zeger (1986) and implemented in the Stata `cluster` option, requires a large number of groups to produce consistent estimates, and thus may not be reliable for the relatively small number of groups (15) in our study.
3. Run regressions (5)-(7) in group averages using weighted least squares (WLS) and use for inferences the critical values of the t -distribution with the number of degrees of freedom equal the number of groups net of the number of regressors (Donald and Lang, 2007). In the presence of regressors varying by individual (such as $y_{i,-1}$ in equation (7)), this solution requires an adjustment as described in Angrist and Pischke (2009, p. 314). More importantly, different weighing schemes in the WLS estimation will produce different point estimates. To keep the WLS points estimates from the regression in averages the same as OLS estimates from the original regression, we will use group size weights.
4. Bootstrap regressions (5)-(7) preserving the within-group correlations in the data. Cameron, Gelbach and Miller (2008) propose a "wild bootstrap" algorithm that does so, which involves replacing the the whole vector of estimated residuals $error_g$ in group g with either $error_g$ or $-error_g$ with equal probability at each bootstrap iteration and reestimating the original equation with the newly generated residuals. We will implement their method.

6 Results

6.1 Baseline results

Table 2 reports the averages and standard deviations of the midterm grade by period (before or after the treatment), group (control or treatment), and status (all, recipients and non-recipients of recognition), as well as estimates of the treatment effect from equations (5) and (6). The treatment effects' standard errors are estimated using simple OLS as well as the four more advanced methods outlined in the previous section. There are signs of a positive effect of recognition on midterm grade, estimated at about 0.5 for the entire sample and 0.6 for non-recipients. Both these estimates are fairly large relative to the grade's standard deviation (about 2.5). However, when we account for the cluster structure of the error term in the underlying regressions (5) and (6) by adjusting the estimates' standard errors we see that they are not precise enough to be considered statistically significant at conventional levels. In the analysis to follow, we will use bootstrap standard errors as the most conservative ones.

[Table 2 here.]

We also observe from the results in Table 2 that there are considerable pre-treatment differences between the control and treatment groups, especially for the non-recipients. These differences go down post-treatment, suggesting that part of the estimated treatment effect may in fact be due to regression to the mean. To account for this possibility, we run the generalized difference-in-difference estimator (equation (7)). The results, reported in Table 3, show that the parameter γ , omitted in equations (5)-(6), is in fact large (about -0.4) and significant, implying that there is indeed convergence between the pre-treatment grade distributions in the control and treatment groups which would have happened even in the absence of treatment. Taking this convergence into account, the treatment effect's estimates are smaller than those previously reported: 0.24 for all, 0.08 for recipients, and 0.35 for non-recipients. None is statistically significant.

[Table 3 here.]

6.2 Treatment effect on other outcomes

Here we analyze the effect of recognition on the final exam grade, presence at the tutorials, the probability of turning up for the midterm, and self-reported hours spent studying the course. All these outcomes are related to student effort and can therefore be affected. Table 4 reports the estimation results from the generalized difference-in-difference equation (7) estimated for each of those variables.⁴

[Table 4 here.]

There seems to be an effect on the recipients' probability of turning up for the midterm, which increases by 0.05 after the treatment. However, with everyone in the control group turning up for the midterm before and after treatment, the treatment effect and its standard deviation are hard to determine. There is also a weakly significant treatment effect on the recipients' self-reported study hours, about a tenth of the hours' standard deviation. But, given the zero correlation between study hours and grade, mentioned earlier, it would be hard to link this result to the rest of our story even if we had no concerns over the reliability of the underlying data.

There is no treatment effect on the final exam grade, which is not surprising because, with the 70% weight in the total grade, its importance is uniformly high for all students regardless of recognition status. Indeed, while the midterm grades' standard deviation is about 2.4, that of the final exam grade is a significantly smaller 1.7. The effect of our rather delicate treatment on the final grade is simply overwhelmed by more important, and uniform, concerns.

Importantly for our further analysis, presence at tutorials is unaffected by treatment. To the extent that this variable is informative of the degree of importance students attach to the course, we can use it to see how it influences the effect of recognition on student performance. Its exogeneity with respect to the treatment allows us to incorporate this variable in our analysis in a simple way as we do in the next section.

⁴Unlike presence at tutorials and midterms, the final exam grade is not available before treatment. We estimate equation (7) for this variable with the first midterm grade added instead.

6.3 The importance of tutorial attendance and effort costs

Whatever theoretical perspective on the effect of recognition on performance, the importance of the task to the agent should strengthen this effect, while higher costs of effort should weaken it. The importance attached to the course is likely to vary by student, and, though not directly observed to us, can be inferred from tutorial attendance. In theory, the frequency of tutorial attendance is an outcome of each student's choosing a utility-maximizing tradeoff between the amount of time spent studying microeconomics and other, competing, activities. Hence, more tutorials attended indicates higher importance attached to the course. In practice, this tradeoff may be affected by the requirement to attend at least 10 out of the total of 13 tutorials in order to be admitted to the final exam and hence to complete the course. This requirement is reflected in the fact that only 2% of the students in our sample attended fewer than 10 tutorials compared to 14% who attended exactly 10 tutorials. A further 20%, 27% and 36% of students attended 11, 12 and 13 tutorials respectively, exceeding the attendance requirement voluntarily.

[Table 5 here.]

The upper part of Table 5 reports estimates of the treatment effect on the midterm grade by the total number of tutorials attended from 10 (the minimum required) to 13 (the maximum possible). These estimates were obtained from the generalized difference-in-difference equation (7) augmented with the interactions of the treatment dummy with the dummy variables corresponding to the number of tutorials attended, which amounts to a triple-difference estimator. We find, as expected, that the treatment effect tends to increase with tutorial attendance, peaking at 0.85 for non-recipients attending 12 out of 13 tutorials (there is still no effect for the recipients). There is a stark increase in the estimated treatment effect from -0.75 for those who attended 10 tutorials, the minimum required to complete the course, to 0.46 for those attending 11 tutorials. The further variation is the treatment effect by attendance is relatively small; in fact, the treatment effects for 11, 12 and 13 tutorial

attendances may be restricted to be the same (the F -statistic for this restriction is 0.26, p -value= 0.77). The single treatment effect for those attending 11, 12 or 13 tutorials is 0.65 with a bootstrapped p -value of 0.08. The latter three results suggest that the 10 attended tutorials is the fault line separating students committed to the course from those merely complying with the attendance requirement.

The results from the same specification where the treatment dummy is interacted with the number of tutorials attended before the treatment, shown in the lower part of Table 5, are smaller in magnitude and statistically insignificant, individually or jointly. The difference between the two sets of results owes itself to the difference in the pattern of attendance before and after the treatment, which is caused in part by attrition. While the majority of students (62%) attended all 8 tutorials before the treatment, only 36% attended all 13 offered throughout the course. Therefore, the students attending 8 tutorials before the treatment are not the same as those attending 12 or 13 in total, hence the blurring of the strong treatment effect we have seen before among the frequent tutorial attendees. Put differently, tutorial attendance before treatment, registered in the beginning of the course, is too noisy a measure of student interest to be useful.

Turning to the effort costs, all else equal, costlier effort should weaken the response to recognition. Although we do not observe them directly, we can infer the costs of effort required to receive recognition from the distance d between the actual grade for the first midterm and the cutoff grade for recognition: the larger d , the costlier the effort. To test whether higher costs of effort to earn recognition reduce the treatment effect for the non-recipients, we allow the treatment effect for this group to vary with d , as well as with tutorial attendance, by interacting the treatment dummy in equation (7) with the quartile of d and with the dummy variable that is 1 when a student attended more than 10 tutorials, and 0 otherwise.

[Table 6 here.]

The results, reported in Table 6, show that the treatment effect increases with both tutorial attendance and proximity to the cutoff grade,

reaching a maximum of 1.32 for the students who attended 11 or more tutorials, thus exceeding the attendance requirement, and were within the first quartile of d ($0 < d \leq 1.5$). Past the first quartile of d , the treatment effect drops substantially in magnitude as well as significance. Controlling for student and tutorial group characteristics (age, gender, group size – the last line in Table 6) produces qualitatively similar results: a treatment effect of 1.10 for the non-recipients who exceeded the attendance requirement and stayed within the first quartile of d , and a much smaller and non-significant effect for the rest of the sample. Though local to a specific group of students in our sample, the treatment effect of 1.10 is substantial and is equivalent to 0.44 of the midterm grade’s standard deviation.

7 Discussion and conclusion

Our analysis has produced three key findings: 1) there is no effect of recognition on its recipients’ further performance; 2) there is a locally positive effect on the non-recipients’ performance; and 3) the effect for the non-recipients increases with tutorial attendance and decreases with the distance to the cutoff grade for recognition. The first two findings are similar to those in Bradler et al. (2013), a study closest to ours, while the third finding, identifying the factors moderating the performance effect of recognition, is completely new. It says that recognition improves student performance when there is enough interest in the activity being recognized and when attaining recognition is not too difficult. Under these conditions, both unaffected by our treatment, the effect is quantitatively large but tails off to insignificance for most of our sample.

Turning to the theoretical explanations to our findings, conformance to the norm is certainly one of those, since this is the only theory which predicts a greater response from non-recipients than recipients, which we observe. It is also most consistent with our experimental setting because it operates with unannounced and cheap to administer recognition, both of which characteristic apply to our treatment. Yet, because conformance also predicts a negative response from the recipients, which we do not observe, it is not the only force behind our results: there must be a

reason for the recipients not to slack. Of the remaining three theories – reciprocity, conditional altruism and status concerns – the latter one is most credible because it does not require recognition to be costly. However, for status concerns to affect effort, recognition must be announced or at least repeated in the future, which was all but excluded by our experimental design (recall the timeline in Figure 1). While expectations of another round of recognition cannot be ruled out completely, the input of the second midterm result in the final grade (10%) may be a more direct reason for the recipients not to rest on their laurels.

We believe our results have important implications for personnel management, generally as well as specific to the business of education. One general implication is that recognition can produce a positive behavioral response even in the presence of other, seemingly more potent incentives such as, in our field experiment, the desire to do well on the course. A further implication, based on the assumption that recognition’s effect is driven by conformance to the norm, is that providing neutral feedback on relative performance would produce the same effect as recognition. A related experimental study by Neckermann and Yang (2012) supports this implication. Therefore, recognition seems to be a popular way of giving performance feedback, containing no useful information in itself.

Specifically to education, the finding that it is non-recipients who respond to our treatment implies that spontaneous recognition of the sort we have implemented may be just the right tool to stimulate the performance of currently underachieving students, as long as they are not too far behind. It may thus be preferable to more formal types of recognition based on tournaments (for example, dean’s lists and other forms of academic distinction) which stimulate to a greater extent the performance of the already more able students. However, care needs to be taken in deciding on the cutoff grade for recognition, since too low a cutoff reduces the information content in recognition, and hence the response to it, while too high a cutoff makes recognition unattainable as well as uninformative.

We conclude by reflecting on the strengths and limitations of our

study, and by stating some outstanding questions that await further research. We believe our experiment is the first to look at recognition in a university environment, a crucial yet under-researched element of modern knowledge economy. Being conducted in a completely natural environment for its participants, and concerning their main activity at the time, it compares favorably in terms of external validity with other studies that feature one-day simple jobs. Yet, the very realism of our experimental design weakens our control over potentially important unobservables, which are all the more important given the heterogeneity in the treatment effect. Does the importance of the group norm vary by group? What is the role of the identity of the recognition giver? Will recognition's effect disappear once students learn their relative performance within the group? Answering these questions through further research will help better understand the workings of recognition and may be useful in maximizing its potential effect.

References

- Akerlof, G.A.** 1982. "Labor contracts as a partial gift exchange." *Quarterly Journal of Economics*, 97(4): 543–569.
- Angrist, J., and J.-S. Pischke.** 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.
- Baron, J., and D. Kreps.** 1999. *Strategic human resources: Frameworks for general managers*. New York: John Willey and Sons.
- Bernheim, D.** 1994. "A Theory of Conformity." *Journal of Political Economy*, 102: 841–877.
- Bradler, S., R. Dur, S. Neckermann, and A. Non.** 2013. "Employee Recognition and Performance: A Field Experiment." Tinbergen Institute Discussion Papers 13-038/VII.
- Chen, Y., F. Harper, J. Konstan, and S. Li.** 2010. "Social comparisons and contributions to online communities: A field experiment on movielens." *American Economic Review*, 100: 1358–1398.
- Donald, S., and K. Lang.** 2007. "Inference with difference-in-difference and other panel data." *The Review of Economics and Statistics*, 89: 221–233.
- Dur, R.** 2009. "Gift Exchange in The Workplace: Money or Attention?" *Journal of the European Economic Association*, 7: 550–560.
- Falk, A., and U. Fischbacher.** 2006. "A theory of reciprocity." *Games and Economic Behavior*, 54: 293–315.
- Grant, A., and F. Gino.** 2010. "A little thanks goes a long way: Explaining why gratitude expressions motivate prosocial behavior." *Journal of Personality and Social Psychology*, 98: 946–955.
- Harrison, G., and J. List.** 2004. "Field experiments." *Journal of Economic Literature*, 42: 1009–1055.

- Hattie, J., and H. Timperley.** 2007. “The power of feedback.” *Review of Educational Research*, 77: 81–112.
- Henderlong, J., and M. Lepper.** 2002. “The effects of praise on children’s intrinsic motivation: A review and synthesis.” *Psychological Bulletin*, 128: 774–795.
- Kosfeld, M., and S. Neckermann.** 2011. “Getting More Work for Nothing? Symbolic Awards and Worker Performance.” *American Economic Journal: Microeconomics*, 3: 86–99.
- Kube, S., M. Marechal, and C. Puppe.** 2012. “The Currency of Reciprocity: Gift Exchange in the Workplace.” *American Economic Review*, 102: 1644–1662.
- Levine, D.** 1998. “Modelling altruism and spitefulness in experiments.” *Review of Economic Dynamics*, 1: 593–622.
- Levitt, S., J. List, S. Neckermann, and S. Sadoff.** 2012. “The Behavioralist Goes to School: Leveraging Behavioral Economics to Improve Educational Performance.” NBER Working Papers 18165.
- Liang, K.-Y., and S. Zeger.** 1986. “Longitudinal data analysis using generalised linear models.” *Biometrika*, 73: 13–22.
- Marzano, R., D. Pickering, and J. Pollock.** 2001. *Classroom Instruction that Works: Research-based Strategies for Increasing Student Achievement*.
- Moldovanu, B., A. Sela, and X. Shi.** 2007. “Contests for Status.” *Journal of Political Economy*, 115: 338–363.
- Moulton, B.** 1986. “Random Group Effects and the Precision of Regression Estimates.” *Journal of Econometrics*, 32: 385–397.
- Neckermann, S., and X. Yang.** 2012. “The impact of scarce rewards on worker effort: An experimental investigation.” Mimeo.
- Stigler, S.** 1997. “Regression towards the mean, historically considered.” *Statistical Methods in Medical Research*, 6: 103–114.

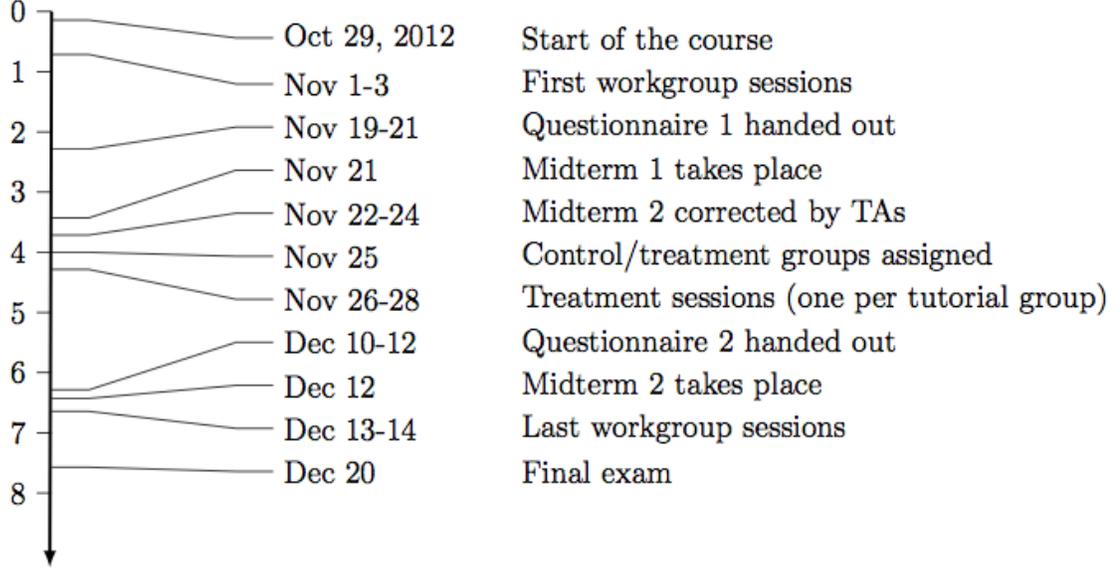


Figure 1: Timeline of the experiment (in weeks)

Table 1: Subjective probability of fulfilling the norm before (π_0) and after (π_1) recognition

	Before recognition $\pi_0(e_i^*, \mathbf{e}_{-i}^*)$	After recognition $\pi_1(e_i^*, \mathbf{e}_{-i}^*)$	After – Before $\pi_1(e_i^*, \mathbf{e}_{-i}^*) - \pi_0(e_i^*, \mathbf{e}_{-i}^*)$
--	--	---	---

The case $\rho \geq \eta$: recognition criterion tougher than the norm

Recipients (R)	$1 - \eta$	1	η
Non-Recipients (NR)	$1 - \eta$	$\frac{(1 - \eta) - (1 - \rho)}{\rho}$	$-\frac{\eta(1 - \rho)}{\rho}$

The case $\rho < \eta$: recognition criterion softer than the norm

R	$1 - \eta$	$\frac{1 - \eta}{1 - \rho}$	$\frac{\rho(1 - \eta)}{1 - \rho}$
NR	$1 - \eta$	0	$-(1 - \eta)$

Table 2: Baseline regression results

	All		Recipients		Non-Recipients	
	Control	Treatment	Control	Treatment	Control	Treatment
Grade before	6.30 (2.38)	5.63 (2.25)	8.44 (1.40)	8.09 (1.04)	5.15 (1.97)	4.35 (1.52)
Grade after	6.11 (2.49)	5.94 (2.47)	7.50 (2.13)	7.36 (1.95)	5.37 (2.36)	5.20 (2.39)
Treatment effect (β_1)		0.49		0.21		0.62
Standard error						
Simple OLS		(0.24)**		(0.36)		(0.29)**
Moulton-corrected		(0.33)		(0.41)		(0.39)
Cluster-adjusted		(0.34)		(0.43)		(0.38)
Group averages		(0.34)		(0.42)		(0.40)
Bootstrap		(0.37)		(0.51)		(0.42)
No. of observations	152	190	53	65	99	125

Notes: Dependent variable: second midterm grade. The OLS estimates of the treatment effect were obtained from equations (5) and (6) run without controls. Standard errors (s.e.) are in parentheses. “Moulton-corrected” s.e. were obtained by applying the Moulton factor to the simple OLS standard errors, using Stata package `moulton` written by Joshua Angrist. “Cluster-adjusted” s.e. were obtained from the residual variance-covariance matrix allowing for within-group residual correlation (the `cluster` option in Stata). “Group averages” s.e. were obtained from the regression in group averages with group sizes used as weights. “Bootstrap” s.e. were obtained from block bootstrapping the original regression equation, using Stata package `bootwildct` written by Bansi Malde. The number of bootstrap repetitions is 1000. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Baseline results allowing for regression to the mean

	All	Recipients	Non-Recipients
Convergence parameter (γ)	-0.37 (0.12)***	-0.39 (0.21)*	-0.34 (0.13)***
Treatment effect (β_1)	0.24 (0.36)	0.08 (0.47)	0.35 (0.44)
No. of observations	342	118	224

Notes: Dependent variable: second midterm grade. Bootstrap s.e. based on 1000 repetitions are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Estimated treatment effects for other outcomes that could be affected by recognition

	All	Recipients	Non-Recipients
Final exam grade			
Treatment effect (β_1)	0.02 (0.25)	0.19 (0.18)	-0.04 (0.46)
No. of observations	336	117	219
Presence at tutorials			
Treatment effect (β_1)	0.06 (0.15)	0.16 (0.24)	0.02 (0.16)
No. of observations	342	118	224
Probability of turning up for the midterm			
Treatment effect (β_1)	0.03 (0.03)	0.05 (0.03)**	0.02 (0.03)
No. of observations	368	121	247
Study hours			
Treatment effect (β_1)	0.85 (1.09)	1.47 (0.80)*	0.36 (1.99)
No. of observations	200	80	120

Notes: Bootstrap s.e. based on 1000 repetitions are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. From this table on, all estimates are based on equation (7) controlling for regression to the mean.

Table 5: Treatment effects by tutorial attendance

	All	Recipients	Non-Recipients
Total tutorials attended [sample frequency]			
10 [15%]	-0.66 (0.74)	-0.42 (0.79)	-0.75 (0.85)
11 [21%]	0.24 (0.43)	-0.57 (0.60)	0.46 (0.58)
12 [28%]	0.56 (0.35)*	0.22 (0.67)	0.85 (0.39)**
13 [36%]	0.32 (0.33)	-0.08 (0.51)	0.55 (0.49)
Treatment effect for 11, 12 or 13 tutorials together	0.44 (0.30)	0.06 (0.42)	0.65 (0.37)*
Tutorials attended before treatment			
5 or 6 [12%]	-0.14 (0.99)	0.24 (1.89)	-0.07 (0.79)
7 [26%]	0.40 (0.39)	0.12 (0.76)	0.48 (0.47)
8 [62%]	0.29 (0.40)	-0.08 (0.39)	0.51 (0.50)

Notes: Dependent variable: second midterm grade. Bootstrap s.e. based on 1000 repetitions are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Treatment effects by tutorial attendance and the distance to cutoff grade (d), non-recipients only

	Quartile of the distance to cutoff grade			
	1st	2nd	3rd	4th
	$0 < d \leq 1.5$	$1.5 < d \leq 2.5$	$2.5 < d \leq 3.5$	$3.5 < d$
All	1.02 (0.54)*	0.46 (0.67)	0.12 (0.74)	0.28 (0.59)
10 tutorials attended	-2.13 (1.74)	-0.56 (1.69)	-0.13 (1.62)	-0.06 (1.45)
>10 tutorials attended	1.32 (0.42)***	0.54 (0.52)	0.48 (0.65)	0.42 (0.86)
The same controlling for group size, study program, age, and gender	1.10 (0.46)**	0.45 (0.41)	-0.35 (0.65)	-0.07 (1.21)

Notes: Dependent variable: second midterm grade. Bootstrap s.e. based on 1000 repetitions are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.