# Recognition in the classroom

Nicky Hoogveld

July 30, 2013

**Abstract**

In this paper we will present an experiment that we carried out with students attending a Microeconomics course at a Dutch university in the autumn of 2012. We found that students who saw their fellow students receiving compliments given by their teacher-assistant, scored almost 0.8 grade points higher on a 1-10 scale for a midterm test when compared with students who were not exposed to compliments. Given the setting of our experiment, our results suggest that norm conformity is among the mechanisms through which recognition impacts performance. Our experiment shows that teachers can improve the study results of their students by giving compliments to good performers.

Master's Thesis Economics of Markets, Organisations & Policy

Supervisor

Nick Zubanov

Erasmus School of Economics

# 1 Introduction

Most basic economic labour theories start from the assumption that workers induce extra effort if they get a monetary compensation for the work they have done. The work by Lazear (2000) indeed shows that this is plausible. However, monetary compensation is not always obtainable or desirable. Luckily for principals the human brain is not only sensible towards money. Psychologists have shown for example that people can be intrinsically motivated and that they are at least partly driven by feelings like pride and shame (Ellingsen & Johannesson, 2008; Bénabou & Tirole, 2005). If employers, or other stakeholders, would be able to explicitly address these kinds of feelings, monetary rewards could be partly replaced by relatively cheap alternatives, like awards or compliments.

So far, there have been a couple of experiments where subjects had to fulfil easy one-time jobs (a.o. Kosfeld & Neckermann, 2011; Bradler et al, 2013), from which the results showed that people are indeed sensible towards compliments. The setting of these experiments was such that most conditions within the experiment could be controlled, which minimised the chance that undesirable factors influenced the outcome of the experiments. The drawback of these experiments is that one could doubt to what extent these 'artificial' tasks can be translated to real world situations. Our experiment tries to fill this gap in the research field.

Our 'natural environment' is the course Microeconomics taught at the Erasmus University Rotterdam during autumn 2012. The students of this course were split into several groups in which they followed workgroup sessions that aimed at preparing them for two midterms and a final exam. Every group was guided by a teacher-assistant who helped the students with homework exercises and who discussed the midterms that have been made by the students. By giving the teacher-assistants the task to either give or don't give compliments to the best performers of the first midterm in their

groups, we were able to find that groups in which compliments were given performed better for the second midterm on average than the groups where this did not happen. This difference was only caused by the improvement of students who scored relatively low for the first midterm. Linking this result with our particular set-up of the experiment suggests that this improvement was triggered by people's desire to conform to norms. Other factors may have contributed as well.

The remainder of this paper will be as follows. Section 2 contains an overview of theoretical and empirical research done in this field so far. In Section 3 a field experiment will be introduced that has been carried out to test to what extent the theories as described in Section 2 hold. Section 4 will analyse the obtained data for this experiment. In Section 5 we will summarise our research and draw our conclusions. The limitations of our research and our recommendations for further research are also discussed in Section 5.

## 2 Literature overview

So far, there has been conducted a lot of research on non-monetary ways to motivate people (a.o. Swank & Visser, 2007; Dur, 2009). One way to do this is by rewarding agents with awards. The experiment by Kosfeld & Neckermann (2011) shows that awards, without any monetary value, can act as motivator for workers. The way they explain this is by implying that actors derive utility from recognition, which can be created by the introduction of awards that are promised to the best workers in advance. Alternatively, awards could increase exerted effort in periods after the award was handed out. Different theories are proposed on the exact mechanism that is behind the motivational power of compliments. Regardless the mechanism, it appears that scarcity should be an element in the design of non-monetary reward systems anyway (a.o. Ellingsen & Johannesson, 2007 and Frey, 2007).

When all peers receive a compliment from their boss, teacher or any other 'credible sender of compliments, this would be seen as simple cheap talk (Farrell & Rabin, 1996). In the remainder of this section we therefore focus on a setting where compliments are only given to subsamples of the entire sample. Besides scarcity, Brophy (1981) argues that compliments must be specific and contingent on the behaviour to be reinforced. In the remainder of this section we will discuss four different mechanisms that could explain why compliments might be a useful way to induce extra effort by workers and how we could be able to distinct between these theories. These are partly taken from Bradler et al (2013). Afterwards we will discuss empirical studies that are related to these theories.

## 2.1 Status

A possible explanation for the motivational power of compliments is that the scarce nature of the reward creates a distinction between receivers and non-receivers. This distinction could be a reason for all subjects to classify their peers and their selves: the students who received compliments can be considered as 'smarter' than their fellow students who did not receive compliments. If we assume that this distinction is indeed creating status differences between those groups of recipients and non-recipients, this would imply that students adjust their level of effort as a result of a change in their utility function. Depending on the distribution of ability of subjects, there exists an optimal number of subcategories to be implemented to get to the highest amount of total effort exerted (Moldovanu et al, 2007). The introduction of compliments as we did in our experiment, in fact generates a situation in which we go from one status category to two status categories: receivers and non-receivers.

Implementing status concerns comes close to the implementation of a tournament system, which could also incentivise subjects (Lazear & Rosen,

1981). Given our specific setting of a public university, we tried to avoid this sort of competition in our experiment: a tournament setting could possibly lead to adverse effects (a.o. Harbring & Irlenbusch, 2008). In the experiment section we will further elaborate on how we did this. Another concern with respect to the design of a status system, is brought forward by Neckermann & Frey (2008): too high requirements for earning a compliment may lead to demotivation among the bad performers.

Although we cannot completely rule out the possibility that status concerns play a role in our experiment, we consider it highly unlikely that these would occur in our setting: status concerns can only be triggered if subjects expect that repetition of the treatment will occur. As a result of the time line of our experiment, the probability that this will indeed happen is very low. Besides, we expect that students do not regard teacher-assistants as great academic authorities. In order words, the value of a compliment given by a teacher-assistant can be considered as close to zero. It is not to be expected that students will proudly tell their friends about the compliment that they received from their teacher-assistant, let alone that they would put it on their resume.

## 2.2 Conformity

A different way in which we could explain performance enhanced behaviour as a result of compliments, is provided by Bernheim (1994). In his model he assumes that people derive utility from the only fact that they appear to be of the type that is considered 'the norm'. In equilibrium there are persons who act differently than they would do naturally. We could adopt this in our experiment in the following way. People can be smart or less smart. A person can only credibly reveal information about his or her smartness by signalling this by scoring grades for exams. A difference with the setting as put forward by Bernheim, is that students do not know exactly what the

academic norm is; simply scoring a grade with which a student passes his exam, might not reflect what the teacher considers as 'good' grades. If a teacher compliments his or her students for scoring high grades, this updates students' believes about the academic norm. Without this compliment, the best guess students could have made about the academic norm was based on all former signals they received. The signalling value of such a compliment depends on at least two factors: the number of signals that have been sent to the receiver in advance and the authority of the sender. Considering the former, the influence of a signal, which a student receives after hundreds of other signals, is smaller than the influence of the first signal in a row. Regarding the latter, it is understandable that students attach less value to a compliment given by an inexperienced teacher than to a compliment given by a teacher who has taught the course for years already. The latter is better able to judge what can be considered as 'the academic norm'.

Bradler et al (2013) explain that conformity can have different implications for receivers and non-receivers. Non-receivers will try to improve to conform to the norm, which will increase their non-material benefits. The receivers may reduce their effort, as a result of a decrease in their belief on the norm. Heterogeneity within the group of receivers and non-receivers could exist: those who score very much below the norm that has been set by the teacher might not even want to try to reach the norm; given their ability, conforming would be too costly.

## 2.3 Gift exchange

The third possible explanation can be found in the literature on gift exchange: people consider it fair to give something in return for a voluntarily given item, regardless its monetary value (Akerlof, 1982). In the long run, gift-exchange can even be explained by game theoretic selfish people: when individuals know that more transactions with their counter party will follow

in the future, it might be rational to 'walk that extra mile' in the current transaction (Rotemberg, 1994). However, in the experiments carried out on gift-exchange and also in our experiment, there is no long-term relationship between parties. The work by Fehr et al (1993) shows for example that, even in a short-term market setting, outcomes can be found which are not in line with standard game theory. When buyers voluntarily offered sellers a larger amount of money for their products, sellers decided to improve the quality of their products. The gift type can largely influence the measured effect (Kube et al, 2012). It is not the monetary value of a gift that determines the reciprocal effect, but rather the time and effort that are required to provide the receiver with the gift.

It is for this reason that gift exchange is a less likely explanation in case we would find an increase in productivity in our experiment. A verbal experiment is almost costless and does not require a significant amount of effort. If gift exchange would be the mechanism, however, it is to be expected that only receivers would increase their effort, as they were the ones who received the gift. According to the theory of gift exchange, the non-receivers would not change their behaviour as a result of the compliments. For them, there is no reason for reciprocal behaviour, because they did not receive anything in the first place.

## 2.4 Signalling

From the perspective of the sender, the compliment can signal that the sender cares about the total population.[1] The sender may give compliments to increase the probability that the population acts altruistically in return for the seemingly altruistic nature of the sender (Levine, 1998). To signal this credibly, the signal needs to be costly (Bradler et al, 2013 and Ellingsen & Johannesson, 2007).

[1]With population we mean both compliment receivers and non-receivers.

When the compliments are given publicly, which is the case in our setting, the positive effect that might be created as a result of this signalling should be homogenously across the population. It should both incentivise the receivers and non-receivers. According to this theory, it is not the compliment in itself that incentives people, but the signal that is sent by means of the compliment.

## 2.5 Empirical evidence

If the positive effect of compliments on future performance is caused by either conformity or status concerns, the compliment can be indirectly considered as the expression of feedback on relative performance. The experiment by Tran & Zeckhauser (2012) about feedback on performance confirms the work by Moldovanu et al (2007). They show that subjects, after they have heard the announcement that the ranking of individual output will be made public, tend to work more productively than they would in absence of this announcement. This is independent from the presence of monetary incentives. Tran & Zeckhauser (2012) even found that people are willing to pay money to get a higher rank and thereby a higher status. The work by Azmat & Iribirri (2010) points in the same direction: their finding is that students of a highschool improved their grades by 5% on average after their quarterly performance report started to include information on their own performance compared with the average performance of their classes. Moreover, their result appeared to be robust across the total sample which is in line with the theory on status concerns. The paper by Blanes i Vidal & Nossol (2011) shows similar results for employees in a firm. Our treatment can be considered as a simplified ranking system with two ranks: students either receive compliments or they do not receive compliments. If status concerns are at stake, this would imply that both recipients and non-recipients would exert extra effort as a result of the introduction of this ranking system. However, a

different effect can be seen in the work by Barankay (2012): in his real world experiment, the performance increased with over 10% after the *cancellation* of a rank order system.

In contrast to conformity and status concerns, the explanations of gift-exchange and signalling have in common that the symbolic value of the publicly given compliment triggers the performance increase. The quasi-experiment by Markham et al (2002) shows an example of such a symbolic reward: privately given compliments about low absenteeism led to less absenteeism in later periods both when compared with the period before compliments were given and when compared with a control group were no compliments were given. Also Kosfeld & Neckermann (2011) measured a positive result when adding non-monetary rewards. The average productivity of students in this one-time job setting appeared to be higher when a symbolic prize was promised to the best performers. In the experiment carried out by Bradler et al (2013) symbolic rewards were given with and without a promise in advance. A productivity increase appeared to be present in the settings where no rewards were promised in advance. Both recipients and non-recipients tended to increase their productivity after the rewards were handed out; only the reaction of the non-recipients is significant at the conventional levels. Our experiment can be analysed in a similar manner: if only non-recipients would react on our treatment, this would give reason to expect that norm conformity is at least one of the mechanisms that would explain the treatment effect.

The longitudinal study presented by Skinner & Belmont (1993) provides an example of gift exchange in the classroom. Pupils tended to show more motivation when they judged their teacher as supportive and structured. The teachers, in turn, increased supportive and structured teacher behaviour when they saw themselves in front of a group of engaged pupils. As this particular study involved a time period of an entire school year, reciprocity

was more likely to occur in this setting than it would be in our experiment: there was only interaction between students and teacher-assistants for about six weeks and the contact frequency was much lower than in a typical class of an elementary school. The longitudinal study by Wentzel (1997) proved the existence of a signalling effect in an educational setting: when students perceived their teachers as caring, this improved their study behaviour. Also in this case the duration differed substantially from the duration of our experiment.

We must also take into account that university students might be less easy to influence with compliments than the elementary and high school pupils in the experiments presented above. Research by Levitt et al (2012) has shown that young children are more responsive towards incentives than older people, especially when it comes to non-monetary incentives.

### 2.6 Summary of theories

The table below, which is partly based on the work by Bradler et al (2013), summarises the theories as discussed in this section.

| Mechanism | Recognition costly? | Repetition required? | Effect on effort | |
|---|---|---|---|---|
| | | | R | N |
| Status concerns | no | yes | + | + |
| Conformance | no | no | - | + |
| Reciprocity | yes | no | + | 0 |
| Signalling | yes | no | + | + |

R stands for receivers, N stands for non-receivers.

Table 1: Summary of theories

# 3 Experiment

## 3.1 Setting

To test which of those explanations is most plausible, an experiment has been conducted with first year students in Economics and Econometrics at the Erasmus University Rotterdam in a natural environment.[2] As part of their first year, students have to follow the course Microeconomics. Their final grade for this course is composed in the following way: during the fourth week of the course the students had to do a test counting for 10%, in week seven they had to do another test counting for 10% and the remaining 80% was determined by their scored average for a weekly internet based test and a final interim exam. Students (N=926 at the beginning of the course) were divided in 29 so-called workgroups where they had the opportunity to practise exam exercises and got additional educational support by teacher-assistants.

Teacher-assistants typically teach one or two workgroups. They have been asked in advance if they wanted to participate in an information session about the experiment that is described below. Due to regulations we were only allowed to ask experienced teacher-assistants[3] to participate in our project. During a session of half an hour we explained the teacher-assistants our experiment and gave them the freedom to leave the experiment if wanted. We explained them that it was crucial that they would not tell anything about this experiment to the students that were part of the experiment. All invited teacher-assistants decided to participate, which gave us a sample size of 601 students divided in seventeen groups.

Additionally, we contacted the teacher-assistants who did not participate in the experiment and asked them how they treated their students. All

[2]The students are not being told that they are part of an experiment.

[3]Teachers-assistants that taught the course Microeconomics for at least one year already.

teacher-assistants responded and described in what way they discussed the midterms. We asked them specifically if they complimented their groups or individuals within their groups. In this way we were able to identify four extra groups (out of the twelve groups that were taught by non-participating teacher-assistants) that were not exposed to compliments. Those groups could therefore act as control groups. This action extended our sample size with 129 students. Students in the remaining twelve groups received some sort of personal compliments[4] by their teacher-assistants and these groups could for this reason neither be classified as treatment groups nor as control groups.

As soon as the participating teacher-assistants had corrected the first set of tests made by their students, the experiment started. After having received the obtained grades, we randomly assigned treatment and control groups. For the treatment groups we additionally calculated what we will call the 'cut-off grade': this is the minimum grade that is scored by the top 30% of each treatment group approximately.[5] We informed the teacher-assistants about the assignment of treatment and control groups. For the treatment groups, we also informed them about the cut-off grade in their group.

The first midterm took place on Wednesday of teaching week four. The session in which the corrected exams were handed out took place on Monday, Tuesday and Wednesday in week five, depending on the teaching schedule of each workgroup.[6] In this session the difference was made between treatment

[4]Either addressed to an individual or to a group of individuals within a workgroup.

[5]Given the facts that students only get grades ending with .0 or .5 and that each group only exists of a finite number of students, it is impossible to set this cut-off point at 30% exactly. For this reason, we chose to set the cut-off point in each group on the grade that approaches this percentage as close as possible.

[6]This time difference allows us to say that the probability of the occurance of an incentive effect is very small: the second test takes places on Wednesday in the last teaching week. Given that each teacher-assistant was told to start discussing the first midterm with their students after almost a week, students probably did not expect that the second midterm would be discussed by their teacher-assistant in the same setting.

and control groups.

After a normal start of the workgroup session in the treatment groups, the teacher-assistant informed the students that their test was going to be handed out and discussed. Thereby he announced the following message in Dutch: "*First, I would like to have your attention for the students to whom I will now hand out their tests, as they did an excellent job. All of them at least scored grade X*[7]*. Experience teaches us that students consider the Microeconomics midterms as very hard. But these students in particular did very well. My compliments!*" While this was said, the tests of the top 30% performers were handed out. The teacher-assistants were told not to further sort the tests. So the order in which the top performing tests were handed out was random. As soon as all these tests had been handed out, the teacher-assistant continued by saying: "*I will now hand out the remaining tests.*" The teacher-assistants handed out the other tests, also in random order. After this, each test question was discussed and after this, the midterms were retrieved again. The session continues in the normal way.

Furthermore, the top tests in the treatment groups were marked with the text "Well done". We asked the teacher-assistants not to include any other tokens of appreciation on question forms (neither in the treatment groups, nor in the control groups), as this could potentially disturb our experiment.

In the control groups, the teacher-assistant also told the students that the midterm would be discussed and handed out. In these groups, however, all tests were handed out in random order without a special message like the one in the treatment groups.

A few days after the grade announcement took place in each workgroup, the grades were published on the information network of the university. Students can learn here how they performed in relation to the entire student population: both the grade distribution is displayed and the percentage of

[7] grade X is the cut-off grade for a particular group.

students that scored the same grade or higher than the logged in student. It needs to be stressed that it does not provide the possibility to see how a student performed compared to his own particular study direction, like Econometrics or Economics. The information value of this publication is therefore limited: students in Econometrics naturally score higher than students in Economics, which could raise the question for students if apples are not being compared with oranges and vice versa. Appendix A shows an example of a potential output a student could get when visiting the information network.

### 3.2 Additional data gathering

To collect additional personal data on students, two questionnaires have been handed out to both students in control and treatment groups. Students have been asked to fill out the surveys for research carried out by a professor. To make sure that students would not link this survey to an experiment, the survey contained several questions that are in no way related to our research. Appendix B shows the full version of both surveys. The questionnaires contained the question to indicate how many hours the students spent on the course in the weeks before and after the first midterm. By asking this question we hoped to gather data on an additional proxy for exerted effort. Our response rate was almost 32%. Unfortunately, we found reasons that made us decide not to use the obtained data on study hours as dependent variable in our regression analysis. The grades scored for the two midterms and the reported number of study hours appeared to be highly uncorrelated: the correlation coefficients were less than 9% and less than 1%, respectively. Taking a closer look at the answers provided to the other questions, we found several questionnaires that were not filled out honestly. Despite the precautions we have taken, we furthermore think that students may have answered some questions strategically, leading to an incorrect estimate of

the number of study hours.

# 4 Results

## 4.1 Data

Our raw data contains 730 observations. The table below shows descriptive statistics. We decided to only consider students who were present during session 8, as this session is crucial in our experiment: in this particular session the difference between control and treatment groups was being made by giving compliments in the last mentioned groups. The sample is futher classified into the different study directions, because it is to be expected that students of a particular study, like Econometrics, have a different ability than other students. We only look at the students of three group types: (Fiscal) Economics, Econometrics and the double-degree Law and Economics students.[8] In the table we do not report students that are so-called "recidivists" and "Leiden-students", as also for further analysis we ignore those students because they are substantially different from other groups.[9] We report grades and presence records in the table, because these indicators are most important for our analysis in the remaining part of this section.

## 4.2 Basic regression

The basic OLS regression that we will estimate for our analysis is as follows:

$$grade = \alpha + \beta_1 * treatment + \beta_2 * time * treatment + \beta_3 * time + \delta * controls + \varepsilon_i$$

[8] Students who study both Economics and Dutch Law are put together in groups.

[9] Recidivists (students who need to redo the course) were put together in a group; these students have no obligatory attendance. This is reflected in attendance reports: many of these students did not show up at all. The Leiden-group is also different from others, because the students in his group got the opportunity to attend longer sessions, because they could not be present during all plenary lectures; on top of that, an extra workgroup session was planned for them.

| | Grade | | Presence | |
|---|---|---|---|---|
| | Control | Treatment | Control | Treatment |
| Before | 6.235 | 5.632 | 0.898 | 0.905 |
| After | 6.032 | 5.937 | 0.853 | 0.862 |
| $N$, of which: | 247 | 190 | 268 | 201 |
| - Economics | 134 | 119 | 146 | 126 |
| - Econometrics | 71 | 71 | 78 | 75 |
| - Mr.drs. | 42 | 0 | 44 | 0 |

Columns for presence show the proportion of attended workgroups.

The $N$ for presence only includes students who were present during both midterms.

Table 2: Descriptive statistics

This difference-in-difference regression (Angrist & Pischke, 2008) enables us to say something about the treatment effect by interpreting coefficient $\beta_2$.

Table 3 shows two specifications where we focus on the grade for the second midterm. Specification II includes attendance dummmies for every session that took place before the second midterm, a female dummy and a dummy for study direction. For Specification I we obtained a significant interaction coefficient at the 10% level of significance. The interaction term for Specification II shows an insignificant interaction coefficient with a p-value of 0.117 (two-sided). Note that the number of observations in slightly smaller when controls are added; the gender information could only be obtained by using the questionnaire results and the student information network, which contains photographs of most students. If both information sources were not available for a student, we could not judge the gender of that particular student. The table suggests that students scored roughly 0.5 grade point higher when they have been exposed to compliments, regardless if they were receiver or witness of classmates receiving compliments. The kernel densities plot for control and treatment groups, as displayed in Figure 1, illustrates this as well: the kernel density plot of the difference between the grades for midterm 2 and midterm 1 for students in the treatment groups is located somewhat right of the plot for students in the control groups.

| | I | II |
|---|---|---|
| | *grade* | *grade* |
| $after$ | -0.158 | -0.128 |
| | (-0.78) | (-0.69) |
| $treatment$ | -0.606* | -0.497* |
| | (-1.74) | (-1.74) |
| $after * treatment$ | 0.527* | 0.441 |
| | (1.78) | (1.65) |
| controls | no | yes |
| *constant* | 6.154*** | 0.0418 |
| | (23.84) | (0.04) |
| $N$ | 899 | 811 |

For this regression table and those presented later in this paper,the following conventions are used:
- $t$ statistics in parentheses.
- $^{*}$ $p < .10$, $^{**}$ $p < .05$, $^{***}$ $p < .01$.
- Controls, if included, are presence for sessions 1-11 (except 8), gender and study direction.
- Standard errors are group clustered.

Table 3: Different specifications for the basic regression

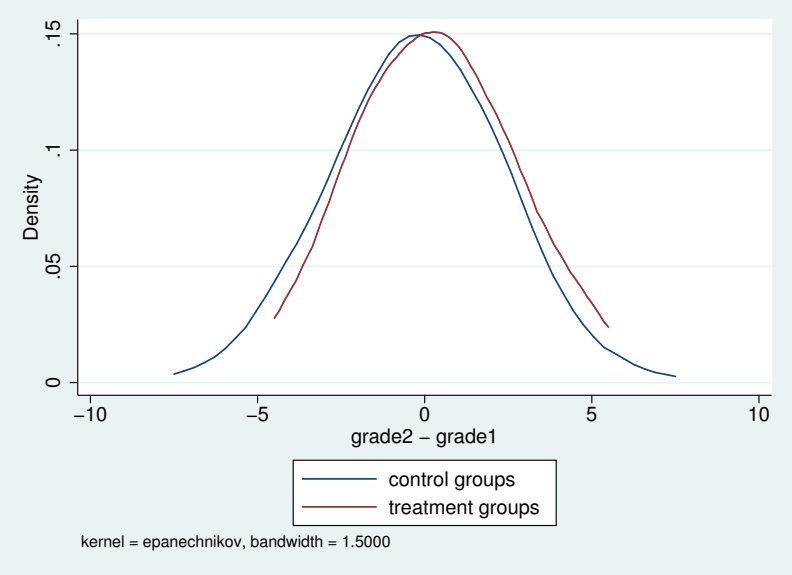



Figure 1: Kernel densities

### 4.3 Recipients and non-recipients

The section above suggests that compliments could help improving students' studying behaviour, it however does not say anything about the potential mechanisms that may be behind this result. In order to link our experiment to the literature that has been presented in Section 2, we need to distinguish students between *recipients* and *non-recipients*. We define recipients as students who approximately belong to the best 30% of their groups. By definition, recipients receive compliments when they are both part of a treatment group and present during session 8. To see what the effect of compliments is on those different groups, we will again estimate the regressions that are shown in Table 3, while now separating between recipients and non-recipients. The result is shown in Table 4. The regression results suggest that the way people responded to compliments depends on the question if they received a compliment or not. For specification I we see that non-receivers scored almost 0.8 grade points higher when they saw that their classmates were complimented for their result. The treatment effect for

| | Non-recipients | | Recipients | |
|---|---|---|---|---|
| | I | II | III | IV |
| | *grade* | *grade* | *grade* | *grade* |
| $after$ | 0.128 | 0.275 | -0.789** | -0.938*** |
| | (0.56) | (1.42) | (-2.59) | (-2.93) |
| $treatment$ | -0.758* | -0.499 | -0.382 | -0.392 |
| | (-2.00) | (-1.46) | (-1.07) | (-1.18) |
| $after * treatment$ | 0.754* | 0.502 | 0.0594 | 0.281 |
| | (2.06) | (1.44) | (0.17) | (0.81) |
| controls | no | yes | no | yes |
| $constant$ | 5.054*** | -0.100 | 8.474*** | 4.903*** |
| | (17.45) | (-0.09) | (27.73) | (4.67) |
| $N$ | 604 | 538 | 295 | 273 |

Table 4: The basic regression for non-recipients and recipients

Specification II is smaller and does not show significance on the conventional levels. The receivers tended to not react at all on compliments: the signs of the interaction term are small and highly insignificant for Specifications III and IV. If we relate these results for receivers and non-receivers to the different theories that are dealt with in Section 2, the theory with which this finding is in line, is the theory of norm conformity. However, it is possible that also additional mechanisms have contributed to this finding.

### 4.4 Regression to the mean

Although Specification I in Table 4 would suggest that the significance of the interaction term shows that compliments indeed lead to better test results, we should be careful with drawing conclusions. When taking a closer look at the summary statistics in Table 2, it can be seen that the average grade for midterm 1 was higher in the control groups than in the treatment groups. Statistical analysis learns us that this difference is significant: when considering all students that were present during session 8, the average differs more than 0.6 grade points with a p-value of 0.098 (two-sided). For this reason,

the significant results presented before could be biased. Assuming that the better result for midterm 1 in the control group was partly caused by luck, it is likely that their difference in grade between midterm 2 and midterm 1 will be relatively worse when compared with the treatment groups. To give an illustration, it is more likely for a student who scored a 3 for his first midterm to improve his grade, than it would be for a student who scored a 9. This phenomenon is called 'regression to the mean' (a.o. Bland & Altman, 1994). As a result of the significant and large difference between control and treatment groups, it is not inconceivable that the specifications in Tables 3 and 4 overestimate the interaction term.

Regression to the mean is also problematic for the reason that the rewarding of compliments to a particular student was dependent on his or her result for the first midterm. Within the groups, compliments are therefore not given randomly. When separating between receivers and non-receivers, this would lead to biased coefficient estimates.

To overcome these problems we continue estimating a dynamic equation, as was also done by Bradler et al, 2013. Our main regression then becomes:

$$grade_2 = \alpha + \beta_1 * grade_1 + \beta_2 * treatment + \delta * controle + \varepsilon_i$$

By using this equation we correct for the grade obtained for the first midterm, which solves the problem of heterogeneity between the randomly assigned control and treatment groups. Table 5 shows several specifications. Specifications I and II are estimated while only considering non-recipients. Specifications III and IV only include recipients. Additionally, Specifications II and IV include controls.

It can be seen that the estimate of the variable *treatment* in Specifications I and II is not significant at any of the conventional levels of significance and that the point estimates are smaller than in the static case. For this reason it can be said that the significant result presented in Table 4 is proba-

| | Non-recipients | | Recipients | |
|---|---|---|---|---|
| | I | II | III | IV |
| | $grade_2$ | $grade_2$ | $grade_2$ | $grade_2$ |
| $grade_1$ | 0.725*** | 0.619*** | 0.697*** | 0.603*** |
| | (9.24) | (6.44) | (3.99) | (3.95) |
| *treatment* | 0.521 | 0.233 | -0.0137 | 0.0224 |
| | (1.45) | (0.86) | (-0.04) | (0.08) |
| controls | no | yes | no | yes |
| *constant* | 1.520*** | -2.791*** | 1.733 | -1.968 |
| | (3.72) | (-3.20) | (1.10) | (-0.97) |
| $N$ | 291 | 259 | 146 | 135 |

Table 5: The dynamic model for non-recipients and recipients

bly at least partly caused by regression to the mean and therefore should be considered as less accurate. The point estimates of the variable *treatment* for recipients are close to zero and far from significant. They confirm what was also shown by the static regressions: it appears that recipients were not influenced by the fact that they have been praised.

### 4.5 Role of commitment

One can imagine that some people are more sensitive towards compliments than others: if a student is in no way interested in a course or in obtaining good study results in general, it might be possible that compliments do not affect that person. This can also be derived from the theoretical framework presented by Bernheim (1994): conformity will only occur if conforming in itself is sufficiently important when compared with intrinsic utility. To check if this is the case, we will run the dynamic equations again, while only including so-called frequent attenders. As we have the presence records for all students, we can use this data as a proxy of commitment; the more often students showed up for their workgroup sessions, the more likely it is that they were interested in the course contents or at least in the final grade that

they would obtain at the end of the course. We decided to mark a student as 'committed' if he or she attended at least eleven out of thirteen sessions for the following reason. Most students[10] were required to attend at least ten out of thirteen sessions: if a student failed to be present at least ten times, the student was not allowed to take the final exam which implies that the course could not be passed. If a student decided to attend eleven or more sessions, at least one of the sessions was visited on a voluntary basis. By only including students who attended at least eleven sessions, we make sure that we only include students who attended the workgroups for a reason that was not related to the fulfilment of this attendance requirement.

Before we run the regression with only frequent attenders, we must be aware of the following. The selection of only frequent attenders may be dangerous for our statistical analysis for two reasons. These reasons relate to the potential influence that treatment may have on presence records of the last five sessions of the course.[11] Firstly, if we would assume that there exists a positive relationship between treatment and presence during the last five sessions, only selecting frequent attenders would lead to a selection of students that are disproportionately receptive towards compliments. If so, this would lead to an overestimation of the treatment effect. Secondly, assuming that treatment indeed influenced presence, the cause-effect relationship between treatment and study results would be unclear. Should we then consider presence as intermediate link of the chain of cause and effect? Or, alternatively, should we consider both variables as independent effects of the treatment that could both act as proxies for exerted effort? To see if this problem occurs, we checked if there is a significant relationship between treatment and presence during the last five sessions. We only con-

[10]Some students are not obliged to attend ten sessions, because they followed the course for the second time. Although most of these students are put in the recidivists group, there are a few exceptions.

[11]These are the sessions that are held after the 8th session, which is the treatment session.

| | $sharepresafter$ |
|---|---|
| $sharepresbefore$ | 0.249* |
| | (2.06) |
| $treatment$ | 0.00694 |
| | (0.24) |
| $constant$ | 0.630*** |
| | (5.79) |
| $N$ | 469 |

Table 6: Effect of treatment on presence after treatment

sider students that were present during session 8. The following regression is estimated:

$$sharepresafter = \alpha + \beta_1 * sharespresbefore + \beta_2 * treatment + \varepsilon_i$$

Variable $sharepresafter$ reflects the percentage of sessions that a student was present after session 8 took place. It is thus calculated by summing up the presence records for session 9 up and until 13, after which it is divided by 5. Variable $sharepresbefore$ reflects the percentage of presence before treatment took place. It includes the treatment session itself, as this is the last session for which we can say for sure that presence is determined exogenously. Table 6 shows the estimation. The coefficient estimate *treatment* is both small and insignificant, which gives us no reason to reject the null hypothesis of no influence of treatment on attendance of sessions that took place after the treatment. The same result is obtained when only considering samples of recipients and non-recipients.

This enables us to run the dynamic specification again while only considering frequent attenders. Table 7 shows the dynamic specification for committed students. Specifications I and II include both recipients and non-recipients; Specifications III and IV include only non-recipients and specifications V and VI are estimated with recipients. Both Specifications

I and II suggest that that an overall treatment effect is present at the conventional significance levels. For non-recipients this effect is even larger: Specification III shows that non-recipients in the treatment groups score almost 0.8 grade points higher for their second midterm than non-recipients in the control groups. The point estimate is smaller when we control for presence per session, gender and study direction, but it is still significant at the 10% level. Specifications V and VI confirm what was also found for recipients in Table 5: recipients seem not to react on the compliments that they have received. The combination of these point estimates and p-values suggests that norm conformity could potentially be the mechanism behind these findings.

| | All students | | Non-recipients | | Recipients | |
|---|---|---|---|---|---|---|
| | I | II | III | IV | V | VI |
| | $grade_2$ | $grade_2$ | $grade_2$ | $grade_2$ | $grade_2$ | $grade_2$ |
| $grade_1$ | 0.621*** | 0.593*** | 0.650*** | 0.629*** | 0.584*** | 0.451*** |
| | (14.16) | (13.56) | (8.22) | (6.26) | (3.51) | (3.07) |
| *treatment* | 0.461* | 0.316* | 0.778** | 0.550* | -0.0861 | -0.193 |
| | (1.88) | (1.83) | (2.55) | (2.00) | (-0.28) | (-0.80) |
| controls | no | yes | no | yes | no | yes |
| *constant* | 2.283*** | -2.474 | 2.010*** | -2.691* | 2.814* | -2.821 |
| | (7.07) | (-1.72) | (4.55) | (-1.81) | (1.90) | (-1.16) |
| $N$ | 358 | 326 | 231 | 207 | 127 | 119 |

Regression only includes students who were present more than ten sessions.

Table 7: The dynamic model estimated with frequent attenders

### 4.6 Heterogeneous effects within the treatment

If the found treatment effect is indeed the result of the tendency of students to conform to the norm, one would expect that students who scored just below the norm would strive to score a better result to conform to the norm next time, whereas students that scored way below the norm would consider themselves unable to conform to the norm, because of a lack of knowledge

or ability. As a result of this, they might be less sensible to norm setting by their teacher-assistant. To check if this is the case we estimate the following model:

$$grade_2 = \alpha + \beta_1 * grade_1 + \beta_2 * treatment + \beta_3 * dif + \beta_4 * dif * treatment + \varepsilon_i$$

The variable $dif$ is the difference between the cut off grade of the group where a student belongs to and the grade actually scored by the student for the first midterm. Table 8 shows specifications for this equation for which we considered only the 'committed' non-recipient students who where present during the treatment session.

| | I | II |
|---|---|---|
| | $grade_2$ | $grade_2$ |
| $grade_1$ | 0.918*** | 0.865*** |
| | (5.76) | (4.20) |
| $treatment$ | 0.999 | 1.038 |
| | (1.70) | (1.56) |
| $dif$ | 0.355 | 0.345 |
| | (1.70) | (1.53) |
| $dif * treatment$ | -0.0190 | -0.173 |
| | (-0.10) | (-0.77) |
| controls | no | yes |
| $constant$ | -0.273 | -5.266*** |
| | (-0.21) | (-2.90) |
| $N$ | 231 | 207 |

Regression only includes students who were present more than ten sessions.

Table 8: Testing for heterogeneous effects

Both specifications show a negative coefficient which is in line with our expectations: with every grade point distance between a student's scored grade and the cut off grade, the treatment effect decreases. However, as these coefficients are not significant, we have to conclude that there is no

significant evidence for the existence of heterogeneity within the treatment.

### 4.7 Further robustness checks

We furthermore ran some regressions to see if gender influences the extent to which productivity changes as an effect of the treatment: in the literature it has been claimed (a.o. Lithari et al, 2010) that females and males react differently on emotional stimuli. As can be seen in Table 9, gender effects were not found in our experiment: the interaction term is highly insignificant. It should be stressed that we can neither confirm nor exclude the possibility that there is an interplay between the gender of the compliment sender (i.e. the teacher-assistant) and the compliment receiver (i.e. the student). Only three groups in the experiment were taught by female teacher-assistants; controlling for this would lead to multicollinearity issues.

| | I | II |
|---|---|---|
| | $grade_2$ | $grade_2$ |
| $grade_1$ | 0.662*** | 0.629*** |
| | (8.68) | (6.28) |
| $treatment$ | 0.602 | 0.571 |
| | (1.50) | (1.51) |
| $female$ | 0.250 | 0.377 |
| | (0.37) | (0.69) |
| $female * treatment$ | 0.108 | -0.0870 |
| | (0.13) | (-0.11) |
| controls | no | yes |
| $constant$ | 2.000*** | -2.729 |
| | (4.52) | (-1.31) |
| $N$ | 207 | 207 |

Regression only includes students who were present more than ten sessions.

Table 9: Testing for gender effects

Starting from Specification IV in Table 7 we changed some control vari-

| | I | II | III | IV |
|---|---|---|---|---|
| | $grade_2$ | $grade_2$ | $grade_2$ | $grade_2$ |
| $grade_1$ | 0.611*** | 0.644*** | 0.397*** | 0.417** |
| | (5.97) | (7.59) | (4.65) | (2.64) |
| $treatment$ | 0.715*** | 0.633* | 0.912* | 0.678 |
| | (3.06) | (1.84) | (2.14) | (1.74) |
| $mrdrs$ | -0.0443 | | | 0.391 |
| | (-0.08) | | | (0.44) |
| $ectrix$ | 0.757*** | | | 1.118*** |
| | (2.92) | | | (3.07) |
| $female$ | | 0.259 | | 0.367 |
| | | (0.62) | | (0.66) |
| $age$ | | | -0.126 | -0.0983 |
| | | | (-1.06) | (-0.95) |
| $constant$ | -2.005 | -2.787 | 5.049 | 4.021 |
| | (-1.33) | (-1.48) | (1.43) | (1.25) |
| $N$ | 231 | 207 | 126 | 119 |

All specifications include presence dummies. Regressions only includes students who were present more than ten sessions and considered as non-recipients. Dummy *lawecon* reflects if students follow the double degree program in Law and Economics. Dummy *ectrix* reflects if student studies Econometrics.

Table 10: Various specifications of the dynamic model

ables to see if the found effect of treatment on performance for the second midterm remains. Table 9 shows the results. The overall message is that the measured treatment effect is robust. The *treatment* coefficient in Specifications I, II and III is significant at the conventional levels.The p-value of *treatment* is just above the 10% level (two-sided) in Specification IV. Note that the number of observations decreases a lot when the *age* dummy is included; information on age could only be derived from returned questionnaires.

# 5 Conclusion, limitations and further research

Our experiment proves that committed students can be motivated by compliments given in a classroom setting. Our OLS estimates show that committed students who were exposed to experiments, either as recipients or as witnesses, scored almost 0.5 grade points higher on a scale from 1 to 10 than their peer students who were not exposed to compliments. When looking at this treatment effect in more detail, it becomes clear that this result was entirely caused by grade improvements made by the students who did not receive compliments. The non-recipients in the treatment groups scored almost 0.8 grade points higher than their colleagues in the control groups. However, the differences between control and treatment groups for the recipients appeared to be small and insignificant. When further focussing on the non-recipient students, we found no evidence that the effect caused by the treatment was heterogeneous across the student population.

When we compare our empirical findings with the literature written in this field so far, the theory of norm conformity as proposed by Bernheim (1994) is compatible with the different effects we find for recipients and non-recipients. The fact that we do not find differences across the group of non-recipients is opposing with his theory: although we found negative coefficient estimates for the treatment dummy, these are far from significant. Also, we cannot exclude the possibility that other, potentially opposing, mechanisms contribute to the total effect we measured.

All in all, based on our experiment, we would recommend teachers to use praise as a way to improve working behaviour by students. This conclusion is in line with most lab setting experiments that were carried out in the field so far. If compliments are 'distributed' as we propose, they do not harm anyone, whereas a significant improvement can be seen in the sub sample of non-recipients.

Naturally, our experiment has its limitations. Our sample size is fair, but

a more robust result would be desirable. By the nature of our experiment we furthermore have to rely on the promise teacher-assistants made saying that they would stay very close to the script we designed. Although we tried to avoid problems with this by giving the participating teacher-assistants the choice to participate in the experiment or not, the field setting of our experiment did not allow us to personally check if all teacher-assistants stuck to the script.

Given the fact that we could only use 21 workgroups in our experiment, it was not possible to implement different designs for the treatment groups. It would have been interesting to see how the treatment effect differs with different levels of compliment scarcity, like it was done in Bradler et al, 2013. The 30 % thumb rule that was used in our experiment was chosen somewhat arbitrary, mostly based on previous literature. Choosing a different cut-off rule could lead to different treatment effects.

As put forward in the literature section, the norm conformity effect is likely to depend on the number of signals that students received before the treatment took place and the academic experience of the compliment sender. As our experiment took place in the first half of a first year undergraduate course, the number of signals that students received on 'the academic norm' is limited. The information value of our treatment is therefore relatively high. An interesting extension of our research would be to carry out a comparable experiment in a later study phase to see if the measured effect indeed decreases as a result of a decrease in information value.

Although Bernheim (1994) explains that norm conformity would suggest the existence of heterogeneity within the group of non-recipients, we could not find this while considering our data on the conventional significance levels. An extension of the model on norm conformity would be desirable to explain this finding.

## References


Akerlof, G. (1982). *Contracts as Partial Gift Exchange*. Quarterly Journal of Economics, Vol. 97, No. 4, pp. 543-569.

Angrist, J.D. & Pischke, J.S. (2008). *Mostly harmless econometrics: An empiricist's companion* (pp. 169-186). Princeton University Press.

Azmat, G. & Iriberri, N. (2010). *The importance of relative performance feedback information: Evidence from a natural experiment using high school students*. Journal of Public Economics, Vol 94, No. 7-8, pp. 435-452.

Barankay, I. (2012). *Rank Incentives. Evidence from a Randomized Workplace Experiment*. University of Pennsylvania.

Bénabou, R. & Tirole, J. (2005). *Incentives and Prosocial Behavior*. NBER Working Paper No. 11535.

Bernheim, B.D. (1994). *A Theory of Conformity*. Journal of Political Economy, Vol. 102, No. 5, pp. 841-877.

Bland, J.M. & Altman, D.G. (1994). *Regression towards the mean*. British Medical Journal, Vol 308(6942), 1499.

Blanes i Vidal, J. & Nossol, M. (2011). *Tournaments Without Prizes: Evidence from Personnel Records*. Management Science, Vol. 57, No. 10, pp. 1721-1736.

Bradler, C., Dur, R., Neckermann, S. & Non, A. (2013). *Employee Recognition and Performance: A Field Experiment*. Tinbergen Institute Discussion Paper 2013-038/VII.

Brophy, J. (1981). *Teacher praise: A functional analysis*. Review of Educational Research, Vol. 51, No. 1, 5-32.

Dur, R. (2009). *Gift Exchange in the Workplace: Money or Attention*. Journal of the European Economic Association, Vol. 7, No. 2-3, pp. 550-560.

Ellingsen, T. & Johannesson, M. (2007). *Paying Respect*. Journal of Economic Perspectives, Vol. 21, No. 4, pp. 135-149.

Ellingsen, T. & Johannesson, M. (2008). *Pride and Prejudice: The Human Side of Incentive Theory*. American Economic Review, Vol. 98, No. 3, pp. 990-1008.

Farrell, J. & Rabin, M. (1996). *Cheap Talk*. Journal of Economic Perspectives, Vol. 10, No. 3, pp. 103-118.

Frey, B. (2007). *Awards as compensation*. European Management Review, Vol. 4, No. 1, pp. 6-14.

Harbring, C. & Irlenbusch, B. (2008). *How many winners are good to have? On tournaments with sabotage*. Journal of Economic Behavior & Organization, Vol. 65, No. 3-4, pp. 682-702.

Kosfeld, M. & Neckermann, S. (2011). *Getting More Work for Nothing? Symbolic Awards and Worker Performance*. American Economic Journal: Microeconomics, Vol. 3, No. 3, pp. 86-99.

Lazear, E. (2000). *Performance Pay and Productivity*. American Economic Review, Vol. 90, No. 5, pp. 1346-1361.

Lazear, E. & Rosen, S. (1981). *Rank-Order Tournamens as Optimum Labor Contracts*. Journal of Political Economy, Vol. 89, No. 5, pp. 841-864.

Levine, D.K. (1998). *Modeling Altruism and Spitefulness in Experiments*. Review of Economic Dynamics, Vol. 1, No. 3, pp. 593-622.

Levitt, S.D., List, J.A., Neckermann, S. & Sadoff, S. (2012). *The Behavioralist Goes to School: Leveraging Behavorial Economics to Improve Education Performance*. NBER Working Paper No. 18165.

Lithari, C., Frantzidis, C.A., Papadelis, C., Vivas, A.B., Klados, M.A., Kourtidou-Papadeli, C., Pappas, C., Ioannides, A.A. & Bamidis, P.D. (2010). *Are females more responsive to emotional stimuli? A neurophysiological study across arousal and valence dimensions.* Brain topography, Vol. 23, No. 1, pp. 27-40.

Markham, S.E., Scott, D. & McKee, G. (2002). *Recognizing good attendance: A longitudinal quasi-experimental field study.* Personnel Psychology, Vol. 55, No. 3, pp. 639-660.

Moldovanu, B., Sela, A. & Shi, X. (2007). *Contests for Status.* Journal of Political Economy, Vol. 115, No. 2, pp. 338-263.

Neckermann, S. & Frey, B. (2008). *Awards as Incentives.* Working Paper No. 334, Institute for Empirical Research in Economics, University of Zurich.

Rotemberg, J. (1994). *Human Relations in the Workplace.* Journal of Political Economy, Vol. 102, No. 4, pp. 684-717.

Skinner, E.A. & Belmont, M.J. (1993). *Motivation in the Classroom: Reciprocal Effects of Teacher Behavior and Student Engagement across the School Year.* Journal of educational psychology, Vol. 85, No. 4, pp. 571-581.

Swank, O. & Visser, B. (2007). *Motivating Through Delegating Tasks or Giving Attention.* Journal of Law Economics and Organization, Vol. 23, No. 3, pp. 731-742.

Wentzel, K.R. (1997). *Student Motivation in Middle School: The Role of Perceived Pedagogical Caring.* Journal of Educational Psychology, Vol. 89, No. 3, pp. 411-419.

# A Output from the university's student information network

Example taken from another course and another year:

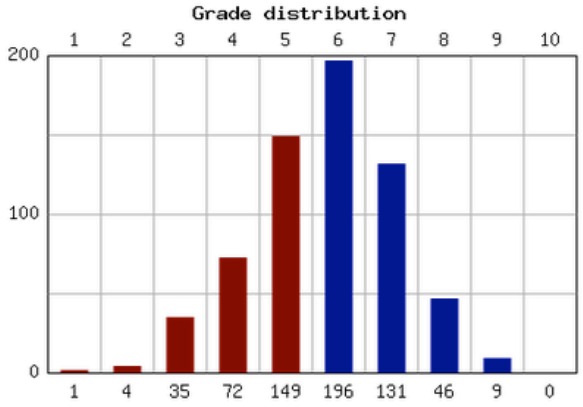
Macro-economie

Grades exam 25-02-2009

| | Numeric results | Alphanumeric results |
|---|---|---|
| Number of results: | 643 | 685 |
| Average: | 5.67 | 5.32 |
| Sufficient: | 382 (59.4%) | 382 (55.8%) |

Your result (8.4) is in the 14 best results.

**You belong to the top 2% (2%) for this exam.**




Not numeric results:

| Value | Count |
|---|---|
| NO | 38 |
| GGR | 4 |

# B Surveys

Survey held in the last session before the first midterm took place:

> Beste student,
>
> Voor research-doeleinden wil ik je vragen deze enquête zo eerlijk mogelijk in te vullen. Met de ingevulde gegevens zal vertrouwelijk worden omgegaan. Ik

wil je erop wijzen dat de op deze manier verkregen informatie op geen enkele wijze gedeeld zal worden met professoren of student-assistenten van het vak micro-economie, noch met andere personen die bij het primaire onderwijsproces betrokken zijn. Ikzelf ben overigens ook niet betrokken bij het onderwijs voor het vak micro-economie. Je studentnummer zal enkel gebruikt worden om de ingevulde informatie te kunnen koppelen aan andere informatie. Voor het verdere onderzoek zal in het geheel met data gewerkt worden die niet op persoonsniveau te herleiden zal zijn.

1. Ben je thuis- of uitwonend?
2. Ben je lid van een studentengezelligheidsvereniging?
3. Hoeveel uren heb je dit blok tot nu toe gemiddeld per week besteed aan het vak micro-economie? (inclusief bezoek hoor-/sommencolleges, werkgroepen en zelfstudie)
4. Op een schaal van 1 t/m 5; hoe goed bevalt het Rotterdamse studentenleven je tot nu toe?
5. Wat is je studentnummer?
6. Wat is je geslacht?

Hartelijk dank voor je medewerking!

N. Zubanov,
Universitair Docent ESE

Survey held in the last session before the second midterm took place:

Beste student,

Een aantal weken geleden heb ik je gevraagd deel te nemen aan een enquête. Voor mijn onderzoek wil ik je vragen opnieuw deel te nemen aan een enquête en deze zo eerlijk mogelijk in te vullen. Opnieuw wil ik erop wijzen dat er met de ingevulde gegevens vertrouwelijk zal worden omgegaan; de op deze manier verkregen informatie zal op geen enkele wijze worden gedeeld met professoren

of student-assistenten van het vak micro-economie, noch met andere personen die bij het primaire onderwijsproces betrokken zijn. Ikzelf ben overigens ook op geen enkele wijze betrokken bij het onderwijs van het vak micro-economie. Je studentnummer zal enkel gebruikt worden om de ingevulde informatie te kunnen koppelen aan andere informatie. Voor het verdere onderzoek zal in het geheel met data gewerkt worden die niet op persoonsniveau te herleiden zal zijn.

1. In welke provincie heb je het grootste gedeelte van je jeugd gewoond? (indien niet Nederland: vul in 'buitenland')
2. Welk(e) profiel(en) heb je voltooid op de middelbare school? (indien niet van toepassing, vul in 'n.v.t.')
3. Hoeveel uren heb je na de eerste tussentoets voor het vak micro-economie gemiddeld per week besteed aan het vak micro-economie? (inclusief bezoek hoor-/sommencolleges, werkgroepen en zelfstudie)
4. Op een schaal van 1 t/m 5; hoe goed bevalt het Rotterdamse studentenleven je tot nu toe?
5. Wat is je studentnummer?
6. Wat is je geboortejaar?

Hartelijk dank voor je medewerking!

N. Zubanov,

Universitair Docent ESE