

TI 2019-082/V
Tinbergen Institute Discussion Paper

Learning from Praise: Evidence from a Field Experiment with Teachers

Maria Cotofan¹

¹ Erasmus University Rotterdam, Tinbergen Institute

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

Contact: discussionpapers@tinbergen.nl

More TI discussion papers can be downloaded at <https://www.tinbergen.nl>

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 598 4580

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900

Learning from Praise: Evidence from a Field Experiment with Teachers*

Maria Cotofan †

Abstract

Non-monetary incentives such as praise are common-place, but their effects on workers performing cognitively-complex tasks remain largely unknown. I expand the teacher incentive literature through a field experiment measuring how repeated public praise for the best teachers impacts teacher performance. Testing different mechanisms, I argue that public praise sends a comparative message, with teachers being motivated when praised and becoming discouraged when not praised. In treated schools, teachers who are unexpectedly praised perform better and teachers who are not perform worse. The positive effect of unexpected praise is persistent and reflects real student learning. The negative effect disappears over time.

JEL codes: C93, I21, J3, J45, J53, M52.

*This paper reports the results from a field experiment for which the design was pre-registered at <https://www.socialscisceregistry.org/trials/2604/history/32360>. I am grateful to all those who provided valuable comments and feedback on early drafts, especially to Robert Dur, Josse Deelfgauw, Bauke Visser, Anne Boring, Jan Stoop, and the participants in many conferences and seminar presentations. This research would not have been possible without the collaboration of 'Adservio Social Innovation SRL' who provided the data and the experimental platform, with a special thanks to Alexandru Holicov and Marian Andrei for their invaluable contributions.

†Corresponding author. Department of Economics, Erasmus University Rotterdam, E Building, office E1-29, 3000DR Rotterdam; Telephone: 0040733680067; E-mail: cotofan@ese.eur.nl

I. Introduction

Non-monetary incentives are playing an increasingly important role in many firms (Gallus and Frey, 2016). From best employee awards and verbal recognition to a sense of identifying with the company's mission, managers can use a broad set of tools to increase the performance of workers. Praise, in particular, now features extensively in popular publications and the business literature, as an effective way to motivate employees (see e.g. Nelson (2012)). However, the effect of praise on effort and performance remains largely unknown. A growing body of experimental research provides evidence for a positive effect of praise on performance (Stajkovic and Luthans, 2003; Grant and Gino, 2010; Kosfeld and Neckermann, 2011; Anderson et al., 2013; Ashraf et al., 2014; Lourenço, 2015; Bradler et al., 2016; Gallus, 2016; Gubler et al., 2016; Hoogveld and Zubanov, 2017). However, the existing evidence is largely limited to short-run effects. Moreover, the evidence is silent when it comes to the effects of repeated praise, is speculative about mechanisms driving such effects, and is confined to jobs involving simple and repetitive tasks. In this paper, I contribute to this body of literature by designing a large-scale field experiment to investigate the long-run effects of praise on performance, and the interplay between announced, un-announced, and repeated praise. I study this question in a setting where employees - 900 teachers in 39 Romanian schools - perform cognitively complex tasks.

There is a growing literature on the effects of providing teacher incentives aimed at improving educational outcomes. However, empirical papers have focused almost exclusively on monetary incentives and have found mixed effects on student performance (Leigh, 2012). Studies in developing countries generally find positive effects of teacher incentives on student test scores and teacher attendance (Glewwe et al., 2010; Muralidharan and Sundararaman, 2011). However, Springer et al. (2011) and Fryer (2013) study large-scale and costly interventions in the US, and find no treatment effects. While it is the fact that providing monetary incentives can increase teacher effort and can lead to better student performance, it can also crowd out teacher intrinsic motivation (Firestone and Pennell, 1993), or lead to cheating or teaching to the test (Holmstrom and Milgrom, 1991; Jacob and Levitt, 2003). What's more, if the incentive scheme is too complex and teachers feel as if they have little control, interventions can have no impact on student achievement (Fryer, 2013).

Little is known, however, about how non-monetary incentives such as public praise impact teacher performance. A number of mechanisms have been put forward to explain why individuals respond to praise in the workplace. First, when praise is provided publicly and only to top performers, it sends a signal about the performance norm at work, such that information about relative performance induces higher effort levels from bottom performers and lower effort levels from top-performers, as both strive to move closer to the apparent performance norm (Bernheim, 1994; Sliwka, 2007; Fischer and Huddart, 2008; Chen et al., 2010; Bradler et al., 2016). Second, when status awards such as praise or job titles are valued and anticipated, they motivate workers to increase effort. Praise activates reputation concerns on the side of the worker, and engages them in a status contest in anticipation of future praise (Moldovanu et al., 2007; Besley and Ghatak, 2008). Third, an agent uninformed of their own ability can get (de)motivated if the principal's actions signal their true ability (Benabou and Tirole, 2003). When effort and ability are complementary, sending a message about relative performance implies a trade-off for the principal between boosting the self-image of some employees, while hurting that of others (Crutzen et al., 2013). This mechanism is also in line with evidence from psychology, on how workers use appraisals as a source of information to gain more accurate self-knowledge (Felson, 1993; Baumeister, 1998).

In this paper, I exploit a dynamic treatment design to shed light on the potential mechanisms driving

teacher responses to public praise. I set-up a randomized intervention in which teachers are praised based on the performance gains of their students. The intervention is repeated at regular time intervals for an entire academic year. In a sample of 900 teachers in 39 Romanian schools, I rank teachers based on improvements in the performance of their students. Teachers are ranked within their own discipline, across all schools. The 25% best teachers within each discipline are labeled as top performers and qualify for praise. I exploit the fact that all schools in the sample use an on-line platform environment to have platform managers publicly praise the top performing teachers in a random half of these schools. In the other half, no praise is provided. Within each school, the platform is regularly used by teachers, students, and parents. While the intervention clearly targets teacher performance, in section VII I further discuss how the treatment could interact with parent and student behavior, and show that this is not likely to influence the results. In treated schools, the intervention gives teachers a very coarse partition of their rank, namely whether they are in the top 25%, or not. In control schools, teachers do not receive any information.

The intervention is repeated twice more in the treated schools, at regular time intervals, throughout the remainder of the academic year. A teacher can be praised repeatedly during the academic year, but teachers can also become top performers for the first time in later rounds. The first intervention is deliberately not announced. During the first intervention it is announced that praise will be repeated, without disclosing an exact date for future rounds. The literature on providing praise distinguishes between an unannounced reward and an announced one. Empirical evidence suggests that announced praise increases the performance of all individuals (Kosfeld and Neckermann, 2011), and has a particularly positive ex-post effect on the performance of non-recipients (Bradler et al., 2016; Hoogveld and Zubanov, 2017). How a combination of the two impacts behavior remains unexplored. The design of my study does not allow for isolating the effect of repeated praise, from that of unannounced praise. However, the experiment I conduct is a significant improvement on the state of the art, because it sheds light on how individuals respond to being repeatedly praised, and it explores how effective the intervention is once they learn to expect it.

The purpose of my experiment is to study the effect of the intervention on student performance gains (based on grades given by the teacher), student attendance, and student performance on anonymously graded standardized exams. My main results are as follows. At the school level, unannounced praise does not have any effect on teacher performance on average. While the average treatment effect at the school level is not statistically significant, there are sizable heterogeneous treatment effects for recipients and non-recipients of praise. Non-praised teachers in the treatment group decrease performance, while praised teachers increase it. The performance of a non-praised teacher in the treatment group decreases by 0.30 standard deviations as compared to similar teachers in the control group. On the other hand, the performance of a praised teacher increases by 0.23 standard deviations as compared to similar teachers in the control group. The effects are large and economically significant.

The treatment response does not vary substantially with the distance from the 25% “top-performer” threshold, confirming that indeed teachers do not know their rank. The results are best explained by a mechanism where praise sends teachers a signal about their performance. As such, updating their beliefs, teachers become more motivated if they receive good news through praise, and become demotivated if they receive bad news through not being praised, in line with the theoretical prediction in Benabou and Tirole (2003) and Crutzen et al. (2013). Repeated interventions do not seem to have any effect on teacher performance. This is true both for teachers who were praised in the past and those who are praised for the first time, suggesting that when teachers learn to expect the reward praise becomes less effective.

Some critics of providing rewards based on performance argue that once incentives or monitoring are conditioned on a performance measure, the said measure ceases to be effective, also known as “Goodhart’s

law” (Goodhart, 1984). For instance, problems arise when the performance measure can be manipulated by employees. Since teachers grade their own students, praise based on the performance gains of their students can incentivise gaming on the side of the teachers. This concern can be addressed when an objective performance measure is available. I use results on high-stake anonymously graded standardized exams, undertaken by final year students. Based on these exam grades, I test whether teachers respond to praise by increasing performance, or if they simply manipulate the performance measure by grading more leniently.

The results indicate that the subjective performance measure does not become a poorer predictor of standardized exam performance in the treated schools. Moreover, I find that positive changes in the performance of students are driven by real learning gains. Praising teachers in the first round raises the grades of their students by 0.17 standard deviations on the anonymously marked exams, undertaken six months after the intervention. The persistence and magnitude of the effect is remarkable, given that final exams cover a broad range of topics. On the other hand, students whose teachers were not praised in the first round do not perform significantly worse on standardized exams as compared to similar students in the control group. Hence, the positive effect that is also observed in the subjective performance measure survives, while the negative effect on subjectively assessed performance disappears over time.

The remainder of this paper is organized as follows. Section II provides an overview of relevant theoretical mechanisms and formulates predictions. Section III introduces the setting, section IV describes the experimental design, Section V presents the results on unannounced praise, and section VI describes the results on announced and repeated praise. Finally, section VII addresses parent and student behavior and section VIII discusses broader implications and concludes.

II. Theoretical Background

Individuals are said to value praise. But why is praise desirable, and what are the underlying mechanisms that drive behavioral responses to praise? Do these channels predict different outcomes for recipients and non-recipients of public praise, and does it matter whether employees expect such an incentive?

There is little to no evidence on the long-run effects of praise,¹ and the existing theoretical mechanisms are limited in predicting behavioral responses to repeated interventions. In this section I review a number of mechanisms that can drive teacher behavior, and discuss the extent to which some of their features apply to my setting, while some dimensions are likely to differ. Specifically, I discuss (i) status contests, (ii) conformity to the norm and (iii) changes in motivation due to learning about performance.

Generally, these mechanisms are integrated in a principal-agent framework. However, for the purpose of this paper, I will focus on the choices of agents. While providing public praise might not always be an optimal strategy for the principal, such considerations are beyond the scope of this paper and I will abstract from them in the remainder of this section.

¹Somewhat related, a number of experimental papers have looked at the effects of public feedback on performance (Blader et al. (2016), Bandiera et al. (2013), Delfgaauw et al. (2013)), with mixed findings. While feedback and praise are closely related, the former is focused on conveying pure relative performance information, while the latter also sends a clear signal of appreciation to workers. Ashraf et al. (2014) disentangle the effects of recognition and feedback, and find that they have opposing effects on performance. While social recognition increases performance, both public and private disclosure of rank information reduce performance.

A. Status contests

A number of studies have shown that agents care about their reputation. This can be driven by the desire to signal a high ability due to career concerns (Bénabou and Tirole, 2006; Swank and Visser, 2006) or because agents wish to be respected by their peers (Grant and Gino, 2010). Praise can send a signal about the quality of a worker. In particular, when praise is given publicly and only to top performers, an element of social comparison is introduced. In my setting, praising top-performers in a way that is visible to colleagues, parents, and students sends a strong signal about the quality of a teacher.

Besley and Ghatak (2008) have postulated that status awards, such as a better job title or calling someone the “employee of the month”, are incentive compatible and they increase effort on the side of the agent while reducing the optimal level of monetary incentives. Moldovanu et al. (2007) predict that agents will seek status awards, as they lead to a higher status within the group. Given the common expectation that praise will be repeated in the future, all agents should increase effort, following the introduction of the reward. Else, in a one-off unannounced intervention, changes in effort due to status concerns should be zero for all agents. To accommodate this mechanism in my setting, during the first intervention all teachers are told that praise will be provided again in the future.

As such, the first hypothesis is:

H1: If teachers' behavior is driven by status concerns, teachers in a treated school will increase performance after the first intervention, independent of whether they were praised or not.

B. Conformity to the norm

Bernheim (1994) argues that when social status is important to individuals, they will conform to social norms. That is because social groups can penalize individuals who deviate from accepted norms, a penalty reflected in a loss of social reputation (Akerlof, 1980). When individuals fear that departures from the social norm will diminish their position within the group, they will conform to a homogeneous standard of behavior, in spite of having heterogeneous underlying preferences.

The provision of public praise sends a signal, to both recipients and non-recipients, about what is the performance norm in the workplace. Individuals who have a preference for conformity will want to adjust their effort such that they are in line with the performance norm (Sliwka, 2007; Fischer and Huddart, 2008; Chen et al., 2010; Bradler et al., 2016). In this case, praise will have opposite effects on the performance of recipients and non-recipients. Those that receive praise learn that they belong to the top 25% of workers, while those who do not receive it learn that they belong to the bottom 75%. If teachers are conformists and like to behave like their peers, then in treated schools top performers should decrease performance, and bottom performers should increase it, so as to get closer to the apparent work norm.

The second hypothesis is:

H2: If teachers' behavior is driven by conformity to the norm, teachers in a treated school will decrease performance if they were praised, and will increase performance if they were not praised.

C. Effect of learning about performance on motivation

Extrinsic and intrinsic motivation have been extensively discussed in the context of incentives, be they monetary or symbolic. Economists generally argue that incentives are a useful tool in promoting effort

and performance, and a significant number of empirical papers support this claim (Gibbons, 1998; Lazear, 2000). However, more recent literature focused on the potential negative spill-overs of such incentives on motivation. The seminal papers of Fehr and Falk (1999), Gneezy and Rustichini (2000b), and Gneezy and Rustichini (2000a) are some of the early examples to report such “hidden costs” of rewards.

Crowding out of intrinsic motivation, in the terminology of Frey (1997), can be a potential mechanism through which rewards reduce the effort of employees. Some lab and field evidence confirms that financial rewards crowd out the intrinsic motivation of agents (Deci, 1971; Kohn, 1999; Fehr and Falk, 1999; Gneezy and Rustichini, 2000b,a). Motivation crowding out appears particularly relevant for public sector employees, such as teachers: a number of studies in the public administration literature and in economics show that public servants tend to be more intrinsically motivated (Buelens and Van den Broeck, 2007; Crewson, 1997; Dohmen and Falk, 2010; Buurman et al., 2012). Georgellis et al. (2010) show that public workers’ motivation can be crowded out by incentives, while Bellé (2015) finds that financial incentives can crowd out the image motivation of workers in jobs with pro-social impact.

To understand why and how rewards impact the motivation of employees, Benabou and Tirole (2003) use the concept of “looking-glass self”, as coined by Cooley (1902). This mechanism postulates that in a principal-agent setting where the principal uses some form of reward or incentive, the agents learn about their own ability through the reward. In other words, such a reward impacts the agent directly through their payoff, and indirectly through their inference process. Benabou and Tirole (2003) argue that empowerment, encouragement, and praise are examples of confidence-enhancement strategies on the side of the principal.

Crutzen et al. (2013) expand on the “looking-glass self” in a setting where ability and effort are complementary in the production function, such that effort levels increase in a workers’ beliefs about their ability. Since agents do not know their ability, they gain self-knowledge and update their beliefs contingent on a comparison signal sent by the employer. The crucial trade-off that the principal faces is between boosting the self-confidence of the best workers, while harming that of the relatively worse ones (Crutzen et al., 2013; Kamphorst and Swank, 2016). As such, in this setting, public praise not only sends a message to the best-performing teachers, but also to those teachers who are not being praised. In other words, if teachers in treated schools learn about their performance through the intervention, then praise sends “good news” about their ability, and not being praised sends “bad news” about their ability.

The third hypothesis is:

H3: If teachers’ behavior is driven by learning about their performance, a teacher in a treated school will become more motivated and increase performance when praised, and will become demotivated and decrease performance when not praised.

D. Predictions

In line with the three mechanisms discussed, the possible treatment effects can be summarized in Table 1. However, it is also possible that several mechanisms play a role at the same time.

Table 1: Treatment effects of public praise according to hypotheses H1, H2, and H3

Recipient	Non-Recipient	Mechanism
+	+	Effect of status incentives (H1)
-	+	Effect of conformity to the norm (H2)
+	-	Effect of learning about performance on motivation (H3)

First, a positive treatment effect for bottom performers in treated schools can be driven by both status incentives and conformity to the norm. If status incentives drive teacher behavior, [Moldovanu et al. \(2007\)](#) predict that all teachers should increase effort proportional to their ability. If the worst performing teachers are also the least able ones, then bottom ranked teachers will increase performance less than those whose performance falls just below the threshold. On the other hand, if a teacher is a conformist, she will increase performance in line with her beliefs about how far below the thresholds her performance falls. As such, a teacher should increase performance more if she believes that she is at the bottom of the distribution, than if she believes that she is ranked just below the performance threshold. However, as teachers do not know their rank, on average all teachers who are not praised should increase performance similarly if their response is driven by conformity.

Second, a positive treatment effect for praised teachers can be explained by both H1 and H3. Repeating the intervention can shed more light on this issue, by looking at those who are praised multiple times. If in the first period teachers learn about relative performance and become more motivated to exert higher effort, repeated praise provides less information. Such a teacher has already learned about their relative performance, and being a top performer again should result in more modest updating²

The repeated provision of praise over a long period of time is an important innovation of this paper. Not only is this, to the best of my knowledge, the first experiment to explore how persistent the effects of praise are over longer periods of time, but it also investigates whether the intervention loses bite once agents get used to the award system. [Rogers and Frey \(2016\)](#) argue that individuals may become desensitized to repeated exposure to a given stimuli. However, in certain instances, repeated interventions can have an effect on behavior. This is the case if the properties of the stimulus are dynamic, or if it is presented at unpredictable intervals. In my setting, repeated praise is announced, but the exact date of the intervention is not known to teachers. Furthermore, different teachers can be praised in repeated rounds, such that subsequent messages still contain a substantial amount of new information. While integrating the short and long run responses to repeated praise in a theoretical framework is beyond the scope of this paper, teacher responses to repeated praise can shed more light on the underlying mechanisms, and can provide useful guidelines for future theoretical and experimental work.

III. Setting

The experiment targets roughly 900 teachers in 39 Romanian schools, who in total have about 20,000 students aged 11 to 18. In Romania, the school year starts in September and continues until the end of June. The education system runs through three 4-year pre-university education cycles: primary school (aged 7-10), secondary school (aged 11-14), and high school (aged 15-18). This experiment will focus on teachers from secondary schools and high schools.

Romania has a centralized education system, and all schools follow the academic curriculum designed by the Ministry of Education. The curriculum provides a detailed guideline of the teaching material. Furthermore, schools use comparable textbooks which are approved by the Ministry of Education, ensuring that teachers use the same materials and proceed with the curriculum in a similar order. As such, schools are homogeneous with respect to the type of information that students learn, and the competencies and skills

²A similar prediction can be derived if the utility from praise is concave in the frequency of praise. Whether this is the driving mechanism can be tested by looking at the response of those who are praised only once. In section VI I show that this mechanism is not likely to drive the results.

they are expected to acquire throughout the school year. My experiment focuses on teachers who teach one of the following nine academic subjects: Romanian language, English language, Mathematics, Physics, Chemistry, Biology, History, Geography, and Computer Science.

At the end of the second and the third cycle (at age 14 and age 18), students are required to undertake standardized national-level examinations in order to graduate from the current cycle, and continue to the next. These standardized exams are high stake, as they help determine high-school and university admission. Undertaken in strictly invigilated exam centres, students work under the supervision of exam inspectors, and class teachers are not present. Exams are graded through a double-blind procedure, by randomly assigned teachers from a different school. As such, class teachers cannot influence their student's performance on these tests by either designing the test, helping students during the examination, or by deciding the grade.

Teachers' wage is independent of their students' performance. The performance of students does not impact teachers' probabilities of promotion either. Teachers typically are subjected to standardized examinations and procedures to earn the right to be hired (*examen de titularizare* for becoming a teacher) or promoted (*gradul didactic I/II*), which do not take student grades into account. As a consequence of that, there is no career incentive for teachers to artificially inflate the grades of students, since they cannot get fired and will not be promoted based on this performance measure. This unique setting allows for cleanly identifying the effect of non-monetary incentives, as teachers cannot leverage praise to gain future monetary benefits.

The format of the academic curriculum is such that each academic year covers new material. For example, while 5th grade students study plant biology, 6th grade students study animal biology, etc. The consequence of this design is that the first grade that the students receive at the beginning of the academic year should reflect the baseline ability of a student, and should be by and large independent from previous learning. As such, I use the first grade that students receive in the beginning of the new academic year, as a proxy for the baseline ability of the student, and in section IV.A I provide additional evidence that this appears to be a reliable measure.

IV. Experimental Design

This experiment follows 39 schools, located in 15 different regions in Romania.³ All the schools in this experiment are making use an on-line education platform which tracks student progress and allows teachers to interact more efficiently with parents. Schools can decide for themselves whether they want to implement the system, and the usage of the platform comes at a small monthly cost.

The platform allows parents to see their children's performance and grade transcripts in real time and easily keep track of their school progress, as teachers regularly post grades and attendance on-line. By working directly with the platform providers and not with individual schools I can ensure that schools, teachers, and parents are not aware of being part of an experiment, and avoid any selection effects into the sample. My experiment thus qualifies as a natural field experiment, following the terminology in [Harrison](#)

³The 39 schools in this experiment perform better than the national average, with average exam grades of 8.48 at the end of secondary school and 8.18 at the end of high school (on a scale from 1 to 10). According to the most recent statistics from the Ministry of Education in Romania (<https://www.edu.ro/rapoarte-publice-periodice>), the average grade for final exams at the country level are 7.44 at the end of secondary school, and 7.83 at the end of high school. However, students from schools in rural areas typically perform worse on the final exams, bringing the national average down. As a result, my sample is roughly representative of schools in the urban area. Since schools are randomly assigned to either treatment or control, quality differences between sampled schools and average Romanian schools are not a threat to the internal validity of this experiment.

and List (2004). Access to the anonymized data allows me to monitor the performance of all students and teachers in the school for an entire academic year.

Schools are randomly assigned to either treatment or control. Teachers at schools which are assigned to the treatment group will receive “public praise”. More precisely, the “best performing teachers” will be publicly praised through a message posted on-line through the management platform. The best performing teachers are those who score among the top 25% across all schools, within their own subject. The first intervention is unannounced. The first intervention announces that public praise will be given again in the future. However, the exact date and frequency of future interventions is not disclosed. Subsequent rounds of praise take place at regular time-intervals until the end of the academic year. Appendix A.1 discusses the experimental time-line in detail.

A. Determining Top Performing Teachers

Building on the literature on teacher productivity, top performers will be determined on the basis of the performance gains of their students. There is by now a fairly large literature on calculating such improvements in student performance due to teacher impact, by looking at teacher value added (Hanushek et al., 1971; Chetty et al., 2014). While it is common in the literature to extract teacher value added from the teacher fixed effect in a regression explaining test score changes (Chetty et al., 2014), this approach is not chosen here for two reasons. First, such an approach might be too hard to explain to the teachers. Fryer (2013) finds that a large scale monetary incentive scheme in New York public schools had no effect on student achievement, despite the intervention totalling a cost of \$75 million. He argues that the most likely explanation for the zero treatment effects was the fact that the scheme was too complex, and provided teachers with too little control. Similarly, teachers might find it difficult to increase performance, when the ranking mechanism is too complex to understand.

Second, calculating teacher value added in the standard way requires that test scores are comparable in terms of content and level, and assumes that students would score similarly across years, in the absence of a teacher effect. In my setting, teachers have some freedom in designing and grading the tests on which the students’ performance gains are calculated, so changes in student performance might not only capture a teacher effect, but also variation in test difficulty over time. To accommodate these two considerations, I calculate performance gains (henceforth PG) using an alternative method to the standard procedure in the literature. While this measure is not directly comparable to the standard teacher value added, in section V.B. I investigate how noisy my measure of performance gains is, and show that it is accurate in predicting student performance on standardized exams, indicating that it is a good measure of student learning.

In my experiment, the school year is divided into four periods. Teacher performance is evaluated for each one of these time periods, namely before each of the three rounds of public praise, and once after the third and final round. Teachers are ranked according to an average of all the individual performance gains of their students, pg_i . Each student’s performance gain pg_i for a period is given by the difference between their baseline performance for the period (denoted by $\theta_{ib,t}$) and their subsequent performance that period $\theta_{i,t}$, where $t \in \{1, 2, 3, 4\}$:

$$pg_{i,t} = \theta_{i,t} - \theta_{ib,t}.$$

$\theta_{i,t}$ is a weighted average of all the subsequent grades of a student that period, where the final grade is

given a weight of 50% and for all other intermediate grades, the remaining weight is equally distributed.⁴ The final grade is given a higher weight because it measures the longest period of time to pass since the baseline performance grade is recorded.⁵

For the repeated rounds ($t=2,3,4$), the performance gain is calculated in a similar manner, where the new baseline performance is replaced by the performance in the previous period, such that:⁶

$$\theta_{ib,t+1} = \theta_{i,t} = pg_{i,t} + \theta_{ib,t}$$

The first grade at the beginning of the school year is used as the baseline performance for the first period, and represents a proxy for student ability. I argue that this is a reliable proxy, as prior to the intervention teachers have no incentive to manipulate grades (teachers are not monitored or rewarded based on student grades). A potential threat to this approach is that a teacher could influence the performance of her students across academic years. For example, a good teacher could put her students on a higher learning path than an average teacher. In that case, students who had good teachers in the past, could have a higher starting baseline performance in the beginning of the new academic year. As a result, the PG of such teachers would be mechanically lower, implying that the best teachers (according to this definition) might not be the ones who are publicly praised.

To investigate whether this is an issue, I make use of a subset of 20 schools (7,742 students and 380 teachers) where data for the previous academic year is also available. Table A.2.1 in Appendix A.2 shows the relationship between current pre-intervention PG, the measure on which teachers are rewarded, and last year's PG. In other words, I test whether a teacher with a high PG in the past is less likely to have a high PG in the current academic year. Even after controlling for a number of student, teacher, and school characteristics, the relationship between performance gains in the previous year and performance gains in the pre-intervention period in the current academic year is weak. Furthermore, the coefficient of interest is positive and rather small (a one standard deviation increase in the previous year's PG translates into a 0.08 standard deviations larger pre-intervention PG in the current year). This suggests that, according to this measure, well performing teachers in the previous year are not less likely to be labelled as top performers in the current year.

In yet another robustness check, I exploit the fact that some students have just joined the school in the beginning of a new cycle, and they have not had the same teacher in the past.⁷ Table A.2.2 in Appendix A.2 shows the relationship between baseline performance and performance gains (column 1) and how this relationship differs by whether a student had the same teacher in the past or not (column 2). As expected, there is a negative relationship between baseline performance and subsequent performance gains: if the baseline performance of a student increases by one standard deviation, the pre-intervention student performance gains decrease by 0.63 standard deviations. This is a mechanical effect, such that if students have a very high starting level, they will naturally have less room for improvement. However, new students do not appear to learn more than recurring ones, nor does the relationship between baseline performance and learning differ

⁴To better illustrate this, take a simple example where at the end of the first period, a student has four grades namely g_1 , g_2 , g_3 and g_4 , in this exact order. Then the initial performance gain $pg_{i,t=1}$ for each student will be calculated using the following formula:

$$pg_{i,t=1} = \theta_{i,t=1} - \theta_{ib,t=1} = \frac{g_2 + g_3}{2} + g_4 - g_1.$$

⁵This weighting method was agreed jointly with educational experts managing the online platform, and it has been pre-registered in the experimental design.

⁶When $\theta_{i,t}$ is missing, $\theta_{i,t-1}$ will be used, and so on. If no previous average exists, $\theta_{ib,t=1}$ is used.

⁷These are students who just started secondary school, or students who just started high-school in schools that don't offer a secondary education program.

across new and recurring students.⁸ This indicates that having the same teacher for two years in a row does not impact learning differently than having a teacher for the first time.

Table 2 presents the average PG per academic subject, across all schools. On average PG are always positive within a subject, varying between 0.09 and 0.25 points. This variation underlines the importance of selecting top performing teachers within their own subject. Specifically, teachers of different subjects most likely require different skills and use different teaching methods, making the comparison between say a math teacher and a history teacher less relevant than between two math teachers or two history teachers.

Table 2: Average performance gains per academic subject, in the beginning of the school year

Subject	Mean	Standard Deviation	No. teachers
Biology	0.253	0.621	63
Chemistry	0.148	0.943	48
Computer Science	0.116	0.712	60
English Language	0.092	0.615	146
Geography	0.249	0.609	65
History	0.139	0.650	65
Mathematics	0.119	0.651	151
Physics	0.247	0.920	84
Romanian Language	0.120	0.614	173

Notes: Columns show the mean and the standard deviation of PG across all subjects, prior to the intervention. PG is expressed in points, and can in principle take any value between -9 and 9.

Teachers' PG each period is defined as an average of all the individual performance gains of their students in that period. A teacher is a top performer if, based on their students' performance gains, they are ranked in the top 25% best performing teachers, within their own subject. Top performing teachers at schools assigned to the treatment group are publicly praised. There are no treated schools, at any point throughout the experiment, in which no teacher is publicly praised. The share of top performers within each school is fairly comparable across schools with a standard deviation of 12.9%.⁹

B. Intervention

After a period of collecting data on teacher and student baseline performance, the first intervention takes place on January 22nd 2018, following the Christmas break. The messages are unannounced and unanticipated. In the schools which were assigned to the treatment group, a message (for the full intervention text, see Appendix A.3) is posted on the front page of the platform, which is visible to all those with a user account (teachers, parents, and students) immediately as they log-in.

The message is addressed to teachers and it states that the platform is interested in how student performance has improved since the beginning of the school year, as it is one of the ways to measure academic progress. The message announces that for a number of academic subjects platform managers have assessed

⁸These results are also robust to estimating the coefficients separately for top performers and for bottom performers.

⁹The perceived scarcity of public praise could also influence how teachers respond to the intervention. Controlling for the share of school-level top-performing teachers within each subject does not change the results either qualitatively or quantitatively, confirming that this variation does not drive the results.

the improvement in student grades across all the schools that implement the electronic platform. Based on this assessment, teachers are informed that a number of teachers in their schools are among the top 25% performers within their subjects, across all the schools in their database. The top performing teachers are listed by name, and thanked for their effort and contribution. Finally, the announcement mentions that such messages will be sent again in the future, to show the platform's gratitude towards teachers' hard work.

The message is highly visible, and seen by all teachers who log-in to post grades or record attendance. To further ensure that all teachers read the message carefully, an additional private message is sent to their personal inbox. The e-mail informs them again about the intervention and provides them with a link to the original post.

The same procedure is repeated twice more throughout the remainder of the academic year, in March and May respectively. Following each intervention, teacher performance is measured on roughly equal intervals of two months.

C. Data and Randomization

Data spans 39 schools from 15 Romanian regions¹⁰ Data collection records the performance of all the students in the school, across the 9 academic subjects of interest. In total, there are 855 teachers¹¹ in the sample, and 19,748 students. Since each student takes on average about 7 of the 9 academic subjects¹² there are in total 130,316 data entries.

Randomization is performed at the level of the treated unit, namely the school, and stratified across three important dimensions:¹³

(i) Student baseline performance : A school-level weighted average of the initial grade that students receive at the beginning of the school year across all subjects, and a proxy for the average student ability in the school.

(ii) Teacher baseline performance: A school-level weighted average of the pre-intervention (since the beginning of the school year) teacher PG, and a proxy for the average teacher quality in the school.

(iii) School size: The number of teachers in the school (who actively use the platform and teach academic subjects).

Due to the limited number of schools, stratification variables are re-coded as binary indicators, as opposed to continuous measures. For example, if the student baseline performance in a school is above the sample average, the binary indicator takes value one, and zero otherwise. Using the three binary indicators, eight strata are constructed. Within each strata, I randomly assign the 39 schools to either treatment or control. Due to a strata with just one school and by splitting the ties in favor of the treatment group, the randomization process assigned 21 schools (55% of teachers in the sample) to the treatment group and 18

¹⁰Some of the schools that use the platform had only recently purchased the rights and were still largely inactive at the time. I drop the schools in which less than 20% of teachers use the platform. In the remaining 39 schools in the sample, 89% of teachers use the platform.

¹¹Some teachers never record any grades in the mentioned period, more precisely 13% of the sample. This indicates some selection with respect to the "type of teacher" that uses the on-line platform. However, this is not a threat to the validity of the experiment: These teachers are similarly distributed between treated and control schools (p-value= 0.455). For the 89% of teachers who use the platform, remaining active did not differ by the treatment status after the intervention, as can be seen in Appendix A.4.

¹²Some subjects are only introduced in later years, and some students only choose, for example, a subset of science subjects.

¹³Together, these three stratification variables capture the main sources of heterogeneity across the 39 schools which are otherwise very similar in the curriculum they use and their administrative structure.

schools (45% of teachers in the sample) to the control group.

Table 3 shows that the randomization process was successful. When comparing schools in the treatment group with schools in the control group, there appear to be no significant differences in terms of either the stratification variables¹⁴ or a number of additional important controls.

Table 3: Balance tests for mean differences between treatment and control

Variable	C	T	P-value
Student baseline performance	7.741 (0.230)	7.873 (0.219)	0.681
Teacher baseline performance	0.142 (0.043)	0.187 (0.045)	0.472
School size (no. teachers)	21.611 (3.240)	22.238 (3.130)	0.890
% Urban schools	0.833 (0.090)	0.810 (0.088)	0.851
% Publicly funded	0.833 (0.090)	0.762 (0.095)	0.594
% Female students	0.524 (0.027)	0.542 (0.023)	0.540
% Male teachers	0.254 (0.060)	0.257 (0.049)	0.279
No. skipped classes	0.745 (0.139)	0.650 (0.105)	0.585
N	18	21	
Multivariate t-test statistics			
F-value	0.233		
P-value	0.982		

*Notes: The first two columns show variable means between the control group of schools, and the treated group of schools. In brackets, standard deviations are presented. The third column shows the p-values from two-sample t-tests on the null hypothesis that group means are equal. Significance levels: *** $p < .01$, ** $p < .05$, * $p < .1$.*

Before the first intervention, roughly 70% of the students for whom a baseline performance measure exists have at least one additional grade. As such, for these students, the PG can be calculated. At the teacher level, this is calculated based on a weighted average of all the individual performance gains of their students, as described in section IV.A. .

Since some students might not have their performance assessed between interventions, the composition of students who determine teacher PG can differ over time. However, the average teacher has 230 students across the multiple classes that they teach, and teachers' PG is calculated based on a substantial share of their students. On average, for each teacher, their PG is calculated based on 112 (54%) students pre-intervention, 160 (72%) students after the first round, 125 (55%) students after the second round and 135 (61%) students in the last round¹⁵. There is no evidence that teachers in treated schools start recording more grades post-

¹⁴To capture potentially fine grained differences, the continuous stratification variables are used in Table 3, as opposed to the binary indicators.

¹⁵Not all grading takes place through a class-level written examination. Students within one class can be graded at different times,

intervention.¹⁶ As such, performance gains are determined by a large number of students for each teacher.

From the 855 active teachers in the pre-intervention sample for whom PG is calculated, for 821 (96%) of them PG is also calculated in the second round of intervention, for 758 (89%) of them PG is also calculated in the third one, and for 729 (85%) of them PG is also calculated in the last round. This attrition is not due to teachers leaving the school, but because none of their students are graded between interventions. Appendix A.4 shows that this attrition does not depend on being assigned to the treatment, on whether a teacher was a top performer or not, nor on the interaction between the two. In total, 56% of teachers qualify for public praise at least once throughout the experiment.

V. Results of unannounced public praise

The effects of public praise on teacher performance are estimated by looking at three outcome variables: (i) PG calculated using class grades given by the teacher, (ii) student attendance, and (iii) standardized exam performance of their students. Data on PG and attendance are collected prior to the intervention, and following each of the three interventions. Standardized exams take place at the end of the school year, for a subset of students ending an academic cycle, aged 14 and 18.

A. Student Performance Gains

To assess the effects of unannounced praise on student performance gains and attendance, I estimate the following two period teacher fixed effects model:

$$Perf_{i,t+1} = \alpha_1 T_{i,t} + \alpha_2 Top_{i,t} + \alpha_3 T_{i,t} * Top_{i,t} + \mu_i + \tau_t + \varepsilon_{i,t} \quad (1)$$

where $Perf_{i,t+1}$ is teacher performance two months after the intervention, measured by either PG or attendance. $T_{i,t}$ is a treatment dummy, indicating whether a teacher was exposed to the treatment or not, such that the treatment dummy takes value 0 for all schools prior to the intervention, and values 0 or 1 after the first intervention, depending on whether the school was assigned to the treatment or the control group. $Top_{i,t}$ is an indicator for being a top performer, namely if a teacher qualifies for being praised at time t+1, by being ranked in the top 25% at time t. $T_{i,t} * Top_{i,t}$ is the interaction between being a top performer and being in a treated school, which takes value 1 for teachers who are publicly praised. μ_i is a teacher-specific fixed effect which captures all time-invariant teacher characteristics, and τ_t is a time fixed effect. The analysis is performed at the teacher level, and the standard errors are clustered at the school level.

Two months after the first intervention, PG is calculated again for 96% of the active teachers in the pre-intervention sample, having a mean of 0.33 points and a standard deviation of 0.62 points (where PG can in principle take any value between -9 and 9, but in sample it ranges between -5 and 5). Table 4 estimates equation (1), using PG as an outcome variable.¹⁷ Appendix A.5 presents the results for attendance.

for example based on class participation.

¹⁶Calculated by looking at the difference in the number of recorded new grades per student after the intervention. The p-value for the coefficient that regresses the number of new grades after the first round on the treatment dummy is 0.686.

¹⁷The standard errors are clustered at the school level. While the number of clusters is larger than the minimally required number of 30, I perform additional robustness checks to exclude the possibility that the standard errors are biased by the fact that there are only 39 schools in the sample. Following [Cameron et al. \(2008\)](#), I implement the wild bootstrap procedure, designed to produce reliable

Table 4: The effect of unannounced public praise on PG

	New PG	New PG
(α_1) Treatment	-0.150 (0.115)	-0.303** (0.121)
(α_2) Top performer		-1.980*** (0.147)
(α_3) Treatment * Top performer		0.528** (0.233)
Teacher Fixed Effects	yes	yes
Time Fixed Effects	yes	yes
N	821	821
F-value	11.75	169.61

Notes: The dependent variable is the PG calculated two months after the first intervention, expressed in standard deviations. In parentheses, heteroskedasticity robust standard errors are presented, clustered at the school level. Significance levels: *** $p < .01$, ** $p < .05$, * $p < .1$.

Column 1 shows that at the school level there is no statistically significant treatment effect of the intervention, although the point-estimate of the average treatment effect is negative. Coefficient α_2 in column 2 reveals that in the control group, teacher performance is in line with mean reversion. If students experience a steep learning curve in the first period, PG will consequently be lower next period, as there is less room for improvement. Reversely, when PG is low in the pre-intervention period, student grades will subsequently increase, as improving is comparatively easy¹⁸

In column 2, treatment effects are separated by whether a teacher was publicly praised or not. In treated schools, those teachers who do not qualify for public praise, but do observe their colleagues being praised decrease performance by 0.30 standard deviations (α_1). On the other hand, those teachers who are top performers and are praised increase performance by 0.23 standard deviations ($\alpha_1 + \alpha_3$), as opposed to top-performing teachers in the control group¹⁹

The behavior of top-performing teachers in treated schools is in line with hypothesis $H3$, as being publicly praised sends teachers a positive message about their performance. Updating their beliefs in response to this signal, praised teachers improve their performance even further. However, a positive treatment effect for praised teachers is also in line with status incentives ($H1$), or a combination between status incentives and learning about relative performance. I further explore this possibility in sections V.B. and VI.A. .

The behavior of the bottom 75% of teachers is also in line with hypothesis $H3$, as both $H1$ and $H2$ predict that following the first intervention, teachers ranked below the threshold in treated schools will increase performance. However, since the bottom category is comparatively large, the crowding out of motivation might not be the same for all the teachers who fall below the threshold in treated schools. For example, teachers who know that their students have made little progress since the beginning of the year could become strongly demotivated if they perceive it to be very difficult to catch up and become top performers in the future. On the other hand, the motivation of teachers who know that their students have improved might be crowded out to a lesser extent, or they could even increase performance, in the hope of being praised in the

standard errors even when the number of clusters is small. The bootstrapped p-values on the coefficients do not change the significance of the results in Table 4, indicating that the number of schools is not a concern for the reliability of the estimated standard errors.

¹⁸Another competing explanation for the observed mean reversion is that PG is in fact a noisy measure of performance. In section V.B. I show that PG has substantial explanatory power for final exam performance.

¹⁹The p-value on the difference between α_1 and $\alpha_1 + \alpha_3$ is p -value = 0.195

future.

To investigate this, I rank the bottom 75% of teachers in three different categories: those in the 1st quantile who qualified for praise (the top 25%), those in the 2nd quantile (between 25% and 50% of the performance distribution), those in the 3rd quantile (between 50% and 75% of the performance distribution), and those in the 4th quantile (the bottom 25% of the distribution) as compared to similar teachers in the control group. The results in Appendix A.6 show that all bottom performing teachers decrease performance following the intervention. Furthermore, teachers in the three bottom quantiles do not respond significantly different to the treatment, indicated by the associated p-values. This confirms that teachers do not have additional information on their relative performance, and only respond to the given binary partition of their rank. In other words, all teachers in treated schools appear to exclusively use the signal that the intervention sends to update their beliefs about their relative performance. In line with the mechanism proposed by Benabou and Tirole (2003) and Crutzen et al. (2013), all bottom-performers seem to become demotivated and perform worse.

B. Standardized exams

To assess the effect of unannounced praise on an objective performance measure which teachers cannot manipulate, I make use of grades on the standardized exam that final year students undertake at the end of the school year. I use a linear regression model which controls for a rich set of student, teacher, and school characteristics. The effect of unannounced praise on exam performance is cleanly identified, because the first round of the intervention is unanticipated and there is no selection into either treatment or into being a top-performer. I estimate the following equation:

$$Grade_{i,j} = \beta_1 T_{i,j} + \beta_2 Top_j + \beta_3 T_{i,j} * Top_j + \beta_4 X_{i,j} + v_{i,j} \quad (2)$$

where $Grade_{i,j}$ is the final exam performance of student i , who has teacher j for that specific exam subject²⁰ $T_{i,j}$ is the treatment dummy and Top_j is an indicator for whether that student's teacher was a top performer prior to the first intervention. The interaction term $T_{i,j} * Top_j$ captures the effect of having a teacher who was praised in the first round, and $X_{i,j}$ is a vector of controls at the student-level (gender, year of study, baseline student performance on the subject, track), teacher-level (subject), and school-level (region, in a rural area, publicly funded, size, baseline student and teacher performance at the school level, and past exam performance of the school). The standard errors $v_{i,j}$ are clustered at the school level.

At the end of the academic year, all year 8 (14 years old) and year 12 (18 years old) students undertake nationally standardized anonymously graded exams. The Ministry of Education makes the exam results publicly available in July. In total, 3,423 students are matched to their exam grade, equivalent to 75% of the total number of final year students in the sample.²¹ The students belong to 335 teachers from the original sample who teach final year classes.

To assess whether the results can be generalized to all students, I test whether final year students system-

²⁰Students take two exams at the age of 14, in Mathematics and Romanian Language. At the age of 18, students take three exams, in Romanian language, a compulsory track-specific subject, and a track-specific subject of their choice.

²¹The matching success rate is contingent on the name of the student being spelled identically in both the platform database, and the database containing the exam grades, provided by the Ministry of Education. For unmatched students, some form of name miss-match occurred. Miss-matches are caused by middle names being omitted in one database, parent initials missing, or varying diacritics. The fully random matching errors do not impact the composition of the sample.

atically differ from the rest of the sample in terms of any observed characteristics. The results in Appendix A.7 shows that final year students do not appear to be different from other students, neither are they more likely to be over-sampled from the treated group. As such, despite only having exam data for final year students, the results are likely to be apply to the whole sample.

I begin by investigating whether pre-intervention PG is a good measure of student learning, as measured by standardized exams performance. Table 5 shows the relationship between the pre-intervention performance gains of each student and their exam performance, controlling for a rich set of student-level, teacher-level and school-level variables.

Table 5: Relationship between pre-intervention PG and exam performance

	Exam grade
Pre-intervention PG	0.317*** (0.030)
Student controls	yes
Teacher controls	yes
School controls	yes
N	5,308
F-value	483.92
R-squared	0.557

*Notes: The dependent variable is the student's exam performance, expressed in standard deviations. The pre-intervention PG is also expressed in standard deviations. OLS regression controls for baseline student performance, student gender, subject fixed effects, class profile fixed effects, degree of urbanization, being a publicly-funded school, school size, baseline student and teacher quality at the school level, past year exam performance at the school level, and region fixed effects. In parentheses, heteroskedasticity robust standard errors are presented, clustered at the school level. Significance levels: *** $p < .01$, ** $p < .05$, * $p < .1$.*

A one standard deviation increase in pre-intervention performance gains translates into a 0.32 standard deviations increase in performance on standardized exams. The coefficient is both statistically and economically significant, and suggests that PG is a good predictor of student learning: even if the pre-intervention performance gain is only based on a subset of the curriculum, it explains a substantial share of the variation in final exams which cover a broad range of topics.

While exam grades are an objective measure of performance, PG could be manipulated by teachers, after the intervention. An extensive cross-disciplinary literature argues that once a measure of performance is used to reward or to monitor, it becomes less effective over time. This has frequently been referred to as Goodhart's law (Goodhart, 1984) and has been linked to the idea of gaming on the side of the agent. In other words, since in treated schools public praise is conditioned on PG, the measure can become less informative about actual performance over time because teachers have the freedom to grade their own students. To address the concern that PG can become a noisy performance measure after the introduction of the reward, I look at the relationship between PG and exam grades, over time. More precisely, I perform a difference in difference analysis on the correlation between the objective performance measure (the standardized exam grade) and the subjective incentivized measure of performance (PG), by estimating:

$$Corr_{i,t} = \lambda_0 + \lambda_1 Post + \lambda_2 T + \lambda_3 Post * T + \lambda_4 X_{i,t} + \varepsilon_{i,t}$$

where the correlation is calculated at the teacher level, across all of their students who undertake the standardized exam. *Post* is an indicator for the post-intervention period, *T* is the treatment dummy, and $X_{i,t}$ is a vector of controls.

The time trend coefficient in Table 6 indicates that PG becomes somewhat less predictive of exam performance over the course of the school year, in all schools. There are no baseline differences between the treated and the control group prior to the intervention, nor does PG become less predictive of exam performance in the treated group. These results clearly indicate that teachers do not attempt to manipulate the incentivized performance measure, once exposed to the treatment.²²

Table 6: DiD analysis on the correlation between PG and standardized exam performance

	Corr(PG, exam)
Post-intervention period	-0.068** (0.030)
Treatment	-0.006 (0.051)
Post-intervention period * Treatment	0.014 (0.055)
Student controls	yes
Teacher controls	yes
School controls	yes
N	497
F-value	17.60
R-squared	0.080

*Notes: The dependent variable is the correlation coefficient, at the teacher level, between the student's exam performance (expressed in standard deviations) and the student's pre-intervention performance gains (expressed in standard deviations). OLS regression controls for average baseline performance of a teacher's students, student gender composition, subject fixed effects, degree of urbanization, being a publicly-funded school, school size, baseline student and teacher quality at the school level, past year exam performance at the school level, and region fixed effects. In parentheses, heteroskedasticity robust standard errors are presented, clustered at the school level. Significance levels: *** $p < .01$, ** $p < .05$, * $p < .1$.*

For 305 teachers who are observed throughout the experiment (and 3,017 of their students), I assess the effect of public praise on their student's exam performance.²³ by estimating equation (2). The results in Table 7 show that the intervention has no statistically significant effect at the school level. The coefficients in the second column are consistent with the effects of unannounced public praise on PG. Those students whose teacher was praised in the first round score 0.17 standard deviations ($\beta_1 + \beta_3$) higher on the final exam in the end of the school year, as compared to their counterparts in the control group whose teacher was a top performer in the first round.²⁴ By comparison, top performers in the control group increase the performance of their students on the final exam by about 0.05 standard deviations (β_2).²⁵

²²This finding is also robust to estimating the equation separately for the sample of top performers, and the sample of bottom performers. The results are available on request.

²³The analysis is performed at the student level. Due to either matching errors, or due to the fact that not all subjects have mandatory exams (for example Science students can choose between Biology, Chemistry, Physics and Computer Science), a small share of teachers have a very small number of students who undertake final exams. As a result of that, their response to the intervention is measured imprecisely, such that giving them equal weight in a teacher-level regressions introduces considerable noise. Teacher level regressions show that all the results are qualitatively the same if one excludes teachers with very few students, or restricts the analysis to subjects which are mandatory.

²⁴The p-value on the difference between β_1 and $\beta_1 + \beta_3$ is p -value = 0.051

²⁵While a high pre-intervention PG is a good predictor of better exam performance, the effect of being a top-performer on exam grades in the control group is a combination of high pre-intervention PG, and strong mean-reversion in the following period, explaining why the β_2 coefficient is comparatively small.

Table 7: The effect of unannounced public praise on standardized exam performance

	Grade	Grade
(β_1) Treatment	-0.055 (0.051)	-0.089 (0.056)
(β_2) Top performer		0.054 (0.052)
(β_3) Treatment * Top performer		0.258*** (0.094)
Student Controls	yes	yes
Teacher controls	yes	yes
School Controls	yes	yes
N	6,639	6,639
F-value	349.27	214.38
R-squared	0.486	0.492

Notes: The dependent variable is the student's exam performance, expressed in standard deviations. OLS regressions control for baseline student performance, student gender, subject fixed effects, profile type fixed effects, degree of urbanization, being a publicly-funded school, school size, baseline student and teacher quality at the school level, past year exam performance at the school level, and region fixed effects. In parentheses, heteroskedasticity robust standard errors are presented, clustered at the school level. Significance levels: *** $p < .01$, ** $p < .05$, * $p < .1$.

In a meta-analysis of studies on interventions in education, [Sanders et al. \(2015\)](#) find that effect sizes are usually no larger than 0.17 standard deviations. The magnitude of the effects on student achievement following a simple non-monetary incentive scheme for teachers are remarkable. By comparison, [Duflo and Hanna \(2005\)](#) find an increase of 0.17 standard deviations in student achievement when teacher's pay is conditioned on their attendance. [Muralidharan and Sundararaman \(2011\)](#) find that incentive pay increased student achievement by 0.12 standard deviations in language and 0.16 standard deviations in math, during the first year.

This indicates that the effects of unannounced praise have persistent effects on student performance, six months after the intervention. The increase in PG due to public praise translates into better performance on final exams, confirming that teachers do not cheat when grading their own students. Rather, the results are in line with real learning on the side of the students, as a result of better teacher performance. Interestingly, the negative effect of not being praised on PG (0.30 standard deviations in Table 4) disappears over time and does not affect final exam performance (0.09 standard deviations in Table 7) in a significant manner.

In section V.A. I have argued that teacher behavior is best explained by hypothesis *H3*. Re-visiting these arguments, the results in Table 7 are also in line with this mechanism. While both status contests and conformity to the norm are incentive compatible with teachers "cheating" when grading their own students, hypothesis *H3* which works through the intrinsic motivation of teachers is not. As the negative effect of not being praised in the first round is phased-out over time, it suggests that in settings where intrinsic motivation is important, in the long run it is easier to motivate than to demotivate employees through public praise.

VI. Results of announced and repeated public praise

A. Student Performance Gains

It is not easy to cleanly identify the effects of repeated interventions on teacher performance. In treated schools, teachers who are praised in the second or in the third round have already been exposed to and affected by unannounced praise. To nevertheless shed some light on the effects of repeated interventions, I undertake an exploratory analysis. I begin by estimating the per-period effect of the combination of unannounced and announced praise on PG and attendance, using the following equation:

$$Perf_{i,t+1} = \sum_{t=1}^t \gamma_{1,t} * T_{i,t} + \sum_{i,t=1}^t \gamma_{2,t} * Top_{i,t} + \sum_{i,t=1}^t \gamma_{3,t} * T_{i,t} * Top_{i,t} + \mu_i + \tau_t + \omega_{i,t} \quad (3)$$

where $t = 1$ is the pre-intervention period, and at $t = \{2, 3, 4\}$ the three intervention rounds take place. As such, at $t = 2$ top performers in treated schools receive unannounced praise. At $t = \{3, 4\}$ top performers in treated schools receive announced praise, and a subset of them receive repeated praise. Standard errors are clustered at the school level.

To explore the effects of repeated public praise, I estimate (1) the effect of not being praised in a given round, (2) the effect of being praised for the first time in any given round, and (2) the effect of being praised repeatedly in any given round, as compared to similar teachers in the control group.

$$Perf_{i,t+1} = \delta_1 T_{i,t} + \sum_{i,j=0}^2 \delta_{2,j} Type_{i,j,t} + \sum_{i,j=0}^2 \delta_{3,j} T_{i,t} * Type_{i,j,t} + \mu_i + \tau_t + \psi_i \quad (4)$$

where $Type_{i,j,t}$ is a categorical variable which records the type j of a teacher i within each period t . Specifically, $Type_{i,j,t}$ takes value 0 if teacher i is not a top performer at time t , value 1 if teacher i is a top performer for the first time at time t , and value 2 if teacher i is a top performer for the second or third time at time t .

Announced (and for some teachers repeated) praise is given two times throughout the remainder of the school year, namely two months and four months after unannounced public praise. After the second round PG is calculated again for 89% of the active teachers in the original sample,²⁶ and after the final round PG is calculated for 85% of the active teachers in the original sample.²⁷ As described in Appendix A.4, this attrition is random and does not relate to being in the treated group, to being a top performer, or to the interaction between the two.

The remainder of this section presents the effects of announced and repeated praise on PG. Appendix A.8 presents the corresponding results for attendance. There are no differences between treatment and control in the number of new grades (per student) that teachers record, confirming that there is no gaming on the side of the teachers at the extensive margin.²⁸ To shed more light on the way learning is distributed throughout the

²⁶Based on roughly 144 students for each teacher, or 61% of all their students, on average.

²⁷Based on roughly 103 students for each teacher, or 44% of all their students, on average.

²⁸Calculated by looking at the difference in the number of recorded new grades per number of students that a teacher has, after each round. The p-value for the coefficient that regresses the number of new grades on the treatment dummy is 0.125 after the second round and 0.947 after the third one.

academic year, Table 8 shows the average PG across treated and control schools, throughout the experiment.

Table 8: PG throughout the intervention

	Treatment	Control	N
Pre-treatment	0.152	0.147	855
Post unannounced praise	0.282	0.387	821
Post announced praise 1	-0.078	-0.032	758
Post announced praise 2	0.473	0.519	729

Notes: Columns show the average PG in the treatment and the control group, throughout the intervention. PG is expressed in points, and can in principle take any value between -9 and 9.

Table 9 provides an overview of the treatment effects across all periods, by estimating equation (3). Announced praise does not seem to have any significant effects on either top or bottom performers, a finding consistent in both repeated interventions. This is in line with the idea that once rewards are anticipated, they tend to lose their effectiveness in moving performance.

Table 9: The effects of unannounced and announced public praise on PG

	New PG
($\gamma_{1,1}$) Treatment Round 1	-0.232** (0.113)
($\gamma_{1,2}$) Treatment Round 2	-0.043 (0.112)
($\gamma_{1,3}$) Treatment Round 3	-0.069 (0.123)
($\gamma_{2,1}$) Top performer Round 1	-0.883*** (0.098)
($\gamma_{2,2}$) Top performer Round 2	-0.760*** (0.133)
($\gamma_{2,3}$) Top performer Round 3	-0.621** (0.236)
($\gamma_{3,1}$) Treatment * Top performer Round 1	0.291** (0.143)
($\gamma_{3,2}$) Treatment * Top performer Round 2	-0.327 (0.195)
($\gamma_{3,3}$) Treatment * Top performer Round 3	0.127 (0.269)
Teacher Fixed Effects	yes
Time Fixed Effects	yes
N	821
F-value	56.47

*Notes: The dependent variable is the PG calculated two months after the previous intervention, expressed in standard deviations. In parentheses, heteroskedasticity robust standard errors are presented, clustered at the school level. Significance levels: *** $p < .01$, ** $p < .05$, * $p < .1$.*

To shed light on the effects of repeated praise on teacher performance, I estimate equation (4) which compares the performance of teachers who were not praised in any given round, with the performance of

teachers who were praised for the first time, and the performance of teachers who were praised for a repeated time within that round. Table 10 presents the results. Praising teachers for the first time (in any round) has no significant effect on their PG in the following period. Since public praise in the first round has a large and significant effect on the performance of both top and bottom performing teachers, the small and insignificant coefficients δ_1 and $\delta_{3,1}$ suggests that the positive effects of public praise observed in Table 4 disappear when teachers anticipate the intervention.

Table 10: The effects of repeated public praise on PG

	New PG	New PG
(δ_1) Treatment	-0.079 (0.086)	-0.115 (0.089)
$(\delta_{2,1})$ Top performer first time		-0.658*** (0.081)
$(\delta_{2,2})$ Top performer repeated time		-1.240*** (0.291)
$(\delta_{3,1})$ Treatment * Top performer first time		0.037 (0.109)
$(\delta_{3,2})$ Treatment * Top performer repeated time		-0.229 (0.353)
Teacher Fixed Effects	yes	yes
Time Fixed Effects	yes	yes
N	821	821
F-value	55.72	43.32

*Notes: The dependent variable is the PG calculated two months after the previous intervention, expressed in standard deviations. In parentheses, heteroskedasticity robust standard errors are presented, clustered at the school level. Significance levels: *** $p < .01$, ** $p < .05$, * $p < .1$.*

Being praised repeatedly has no significant effect on teacher performance, and the point-estimate is negative. This is in line with hypothesis $H3$, as teachers who were already praised once in the past update their beliefs to a much lesser extent when receiving an additional positive signal. An alternative explanation for coefficient $\delta_{3,2}$ is that the utility from praise is concave in the frequency of praise. However, concavity implies that those teachers who are praised only once should increase performance in all rounds. The much smaller size of coefficient $\delta_{3,1}$ in Table 10, as compared to coefficient α_3 in Table 4, suggests that this is not a plausible mechanism.

Equation (4) imposes the restrictive assumption that in treated schools the response of any teacher i at time t is independent of their experiences in previous rounds. However, repeating the intervention in treated schools and exposing the same group of teachers to multiple treatments gives rise to increasingly complex combinations of effects with each additional round. To relax this assumption, in Table 11 I estimate a flexible specification controlling for each type of experience that a teacher could have had throughout the year, such that at each point in time the previous performance of a teacher is taken into account. The reference category is made up of teachers who were never praised, up to that period.

Table 11: Treatment effects throughout all periods

	PG 1 round	PG 2 rounds	PG 3 rounds
(ψ_1) Treatment	-0.303** (0.121)	-0.249** (0.099)	-0.265** (0.104)
Treatment * Type			
(ψ_2) T*Top1	0.528** (0.233)	0.473** (0.215)	0.490** (0.221)
(ψ_3) T*(NTop2 & Top1)		0.298 (0.279)	0.314 (0.278)
(ψ_4) T*(Top2 & NTop1)		-0.124 (0.524)	-0.116 (0.254)
(ψ_5) T*(Top2 & Top1)		-0.068 (0.445)	-0.051 (0.449)
(ψ_6) T*(NTop3 & NTop2 & Top1)			0.525* (0.301)
(ψ_7) T*(NTop3 & Top2 & NTop1)			0.179 (0.261)
(ψ_8) T*(NTop3 & Top2 & Top1)			0.401 (0.483)
(ψ_9) T*(Top3 & NTop2 & NTop1)			0.106 (0.196)
(ψ_{10}) T*(Top3 & NTop2 & Top1)			0.606 (0.536)
(ψ_{11}) T*(Top3 & Top2 & NTop1)			-0.625 (0.490)
(ψ_{12}) T*(Top3 & Top2 & Top1)			-0.455 (1.028)
Type Fixed Effects	yes	yes	yes
Teacher Fixed Effects	yes	yes	yes
Time Fixed Effects	yes	yes	yes
N	821	821	821
F-value	169.61	88.53	122.81

Notes: The dependent variable is the PG calculated two months after the previous intervention, expressed in standard deviations. In parentheses, heteroskedasticity robust standard errors are presented, clustered at the school level. Significance levels: *** $p < .01$, ** $p < .05$, * $p < .1$.

While the effects of unannounced praise remain sizable and precisely estimated, repeated interventions do not appear to significantly affect teacher performance. However, some interesting patterns arise. First, teachers who are praised for the first time in subsequent rounds (ψ_4 and ψ_9) do not improve performance. Second, those teachers who were only praised in the first round and became more motivated as a result of that, appear to exert additional effort to maintain a high performance throughout (ψ_3 and ψ_6). This is particularly visible following the final intervention, with a marginally significant increase in PG, as compared to similar teachers in the control group. Third, there are no benefits from praising a teacher in two consecutive interventions (ψ_5 , ψ_{11} and ψ_{12}).

These results suggest a number of additional takeaways. When rewards are given repeatedly, being first to the prize seems to matter more. Teachers who are praised in the first round increase performance (ψ_2) and

appear to remain more intrinsically motivated throughout the remainder of the experiment (ψ_3 and ψ_6). On the other hand, being second or third to the prize does not translate into better performance. Finally, repeated rewards over short periods of time do not achieve the desired results, as coefficients on being praised two or three periods consecutively are always negative (ψ_5 , ψ_{11} and ψ_{12}). However, being praised in the first round and in the third one returns a large and positive coefficient (ψ_{10}) following the final intervention. This indicates that while praise loses bite as it becomes less scarce, it remains a powerful tool for those who receive it sparingly.

B. Standardized exams

For the effect of repeated public praise on final exam performance, I estimate:

$$Grade_{i,j} = \phi_1 T_{i,j} + \psi_2 Freq_j + \psi_3 T_{i,j} * Freq_j + \psi_4 X_{i,j} + v_{i,j} \quad (5)$$

where $Grade_{i,j}$ is the final exam performance of student i , under teacher j . $Freq_j$ is a categorical dummy which indicates whether, and how many times, a teacher was a top performer throughout the academic year. In other words, $Freq_j$ takes value 0 if a teacher was always a bottom performer, value 1 if a teacher was a top performer only once, and value 2 if a teacher was a top performer repeatedly throughout the year. The vector of controls $X_{i,j}$ is defined as in equation (2), and standard errors $v_{i,j}$ are clustered at the school level.

In Table 12 I undertake an exploratory analysis on the effects of repeated praise on final exam performance. Specifically, I estimate equation (5) which classifies teachers as always bottom performers (135 teachers and 2,192 students), top performers only once throughout the experiment (138 teachers and 2,281 students) and top performers more than once (45 teachers and 1,027 students).

Dis-aggregating the results by the frequency with which a teacher was a top performer reveals some heterogeneity. The students of those teachers who were never praised do not perform any different than their peers in the control group, with a point estimate of zero. The students of teachers who were only praised once throughout the academic year do not perform significantly different either. On the other hand, the students of repeatedly praised teachers perform 0.33 standard deviations better on final exams, as compared to their counterparts in the control group.²⁹ This effect appears to be predominantly driven by teachers who were praised in the first and in the third round (roughly 60% of the teachers who were praised repeatedly). In line with coefficient ψ_{10} in Table 11, these findings suggest that in the long-run, repeated public praise can be an effective tool if given sparingly. However, those teachers who are top performers multiple times in the treated group are a select type, and have predominantly also been praised in the first round. Thus, the results should be interpreted accordingly and with caution.

²⁹The p-value on the difference between ψ_1 and $\psi_1 + \psi_{3,2}$ is $p\text{-value} = 0.000$

Table 12: The effect of repeated public praise on standardized exam performance

	Grade	Grade
(ϕ_1) Treatment	-0.055 (0.051)	0.006 (0.079)
$(\phi_{2,1})$ Top performer only once		0.132 (0.132)
$(\phi_{2,2})$ Top performer more than once		0.038 (0.052)
$(\phi_{3,1})$ Treatment * Top performer only once		-0.176 (0.144)
$(\phi_{3,2})$ Treatment * Top performer more than once		0.338*** (0.082)
Student Controls	yes	yes
School Controls	yes	yes
N	6,639	6,639
F-value	349.27	587.47
R-squared	0.486	0.493

*Notes: The dependent variable is student exam performance, expressed in standard deviations. OLS regressions control for baseline student performance, student gender, subject fixed effects, class profile fixed effects, degree of urbanization, being a publicly-funded school, school size, baseline student and teacher quality at the school level, past year exam performance at the school level, and region fixed effects. In parentheses, heteroskedasticity robust standard errors are presented, clustered at the school level. Significance levels: *** $p < .01$, ** $p < .05$, * $p < .1$.*

VII. Teacher and parent response

Publicly praising top performers in this experiment implies that parents and students also learn about the quality of a teacher. Thus, it is possible that they adjust their behavior as a response to the intervention. Parents learning about the quality of teachers could partly drive the treatment effect if the amount of time spent home with children changes due to the message (Pop-Eleches and Urquiola, 2013). In other words, parents whose children’s teacher is revealed to be a bottom performer could increase the amount of time they spend helping them with homework, and reduce effort if they find out that the teacher is particularly good. Student behavior could also be affected by the intervention through both a direct channel, and indirectly through teacher effort. In other words, the student’s optimal choice of effort could also be affected by the signal their teacher receives.

However, if praising a teacher sends a clear positive signal about their performance, not praising a teacher sends a much more ambiguous message to teachers, students, and parents; and while the intervention provides an explicit incentive for teachers, for students and parents it is at most an information treatment. For example, while teachers internalize the reward and stop responding over time, there is less reason for students to stop responding when their teacher is praised for the first time, in a later round.

To nevertheless address these concerns, I perform an additional robustness check to investigate whether parents and students adjust their behavior as a response to the intervention. Specifically, I focus on one decision that students or parents can make as a response to the treatment, by looking at the subset of students finishing high school and undertaking the exit standardized exam at the end of the school year (*examenul de bacalaureat*). Depending on the class track, these students undertake two compulsory exams on track-

specific subjects, and additionally select one subject of their choice for their third exam.³⁰ Students must officially choose the elective discipline and register for the exam in the end of May, roughly one month before the exam, and five months after the first intervention.

Using the Ministry of Education track-specific guidelines, I restrict the analysis to exam grades for elective subjects. I then ask whether a student is more likely to choose to undertake an exam on a subject where the class teacher was unexpectedly praised, or less likely to do so if they learn that their teacher is a bottom performer. This would indicate that students or parents directly respond to the intervention, by adjusting educational decisions based on the signal sent by public praise.

Results in Appendix A.9 show that students are not more likely to select an exam subject once they learn that the teacher is particularly good, neither are they more likely to avoid a subject when they learn that the teacher is a bottom performer. This appears to indicate that parents and students do not adjust educational decisions after learning about the performance of teachers. While the results in Appendix A.9 only look at a subset of the total sample (137 teachers teaching elective subjects in the final year to 1,990 students), they provide some evidence that indeed the conclusions of this paper are driven by teacher behavior.

VIII. Conclusion

This paper has shown that introducing public praise as an incentive can have large and persistent effects on the performance of teachers. Unlike experiments studying the short-run effects of praise in simple jobs, this is the first study to assess the how persistent the effects of public praise are and to measure the effects of praise on employees who perform cognitively complex tasks. By analyzing the interplay between unannounced and announced praise, this experiment exploits a dynamic treatment design to shed light on the different theoretical mechanisms that drive teacher behavior in this setting.

Unannounced praise has large effects on teacher performance, consistent with a theoretical mechanism where teachers do not know their performance, but learn about it through public praise. In treated schools, being praised sends a positive signal to top performing teachers, while not being praised sends a negative signal to bottom-performing teachers. Receiving good news boosts confidence and increases teacher performance, while bad news demotivates and reduces teacher performance. The findings are in line with the comparative cheap talk literature, where principals face a trade-off between boosting the confidence of the best performing agents, while harming that of the worst performing ones. A public message where praise is scarce appears to be credible, as teachers seem to update their beliefs accordingly.

Conformity to the norm does not seem to have a bite in this setting. Since teachers are compared to both peers from within their own school, and peers from different schools, the pressure to conform might be reduced. Top performing teachers can more easily increase performance without fearing social punishment, while those at the bottom feel less pressure as the message does not reveal how poor their performance is within their own school. Future experimental work could vary the salience of workplace social norms to further explore this channel.

The behavior of bottom performing teachers does not seem to be driven by status contests either. This could be partly explained by the fact that teachers are intrinsically motivated public servants, such that the

³⁰For example, students in the science track can choose one discipline between Biology, Chemistry, Physics and Computer science, but exams in Mathematics and Romanian Language are compulsory.

utility from winning a status contest is likely outweighed by crowding out of intrinsic motivation when not rewarded. The fact that teachers do not respond to praise in repeated rounds, even when experiencing it for the first time, is in line with the idea that once rewards for a certain performance measure are internalized, expected, and perceived as less scarce, they tend to lose bite. This suggests that the marginal benefits to continued treatment are small even when the composition of top-performing teachers changes over time.

The findings of this paper do not support the idea that teachers attempt to manipulate performance gains. This is to be expected if teacher behavior is driven by an intrinsic channel (*H3*), as opposed to extrinsic motivators (*H1* and *H2*). Unannounced praise leads to real learning gains and better performance on anonymously graded standardized exams for the students of top performing teachers. By the end of the school year, the negative effect of unannounced praise due to motivation crowding out does not reflect the exam performance of students whose teachers were not praised. The fact that the negative responses are short lived suggests that when employees are sufficiently intrinsically motivated, in the long-run it is easier to motivate, than to demotivate through public praise.

An extensive literature has discussed the complexity of rewarding teachers: incentive based subjective assessments can lead to gaming, or teaching to the test. Monetary incentives in such settings are costly and often inefficient (Goodman and Turner, 2013; Eberts et al., 2000). Reasons usually invoked are the complexity of the incentive scheme, or the fact that incentives are too small (Fryer, 2013). In particular, small monetary incentives can backfire in pro-social jobs as the motivation of workers is crowded out. A simple system of non-monetary rewards could be a cheap and efficient alternative, as the positive effect of unannounced praise is persistent, while the negative effect fades away as teachers who are demotivated at first appear to overcompensate in the long-run.

The fact that repeated public praise is ineffective suggests that managers should not provide the same non-monetary reward too often- but rather be creative in varying the incentive, as well as credibly commit to a public and comparative message. To limit the negative effect on bottom performers, praise should have a positive focus, and implicit shaming should be avoided by keeping the bottom category broad. Finally, much more can be said about the different settings in which these incentives are effective, contingent on the type of task and the preferences of the worker. Focus has recently shifted on the reduced importance of income, as many modern workers are increasingly in search of non-monetary work aspects such as meaningful work, flexibility, and appreciation (Cassar and Meier, 2018). As the nature of work changes, better understanding such rewards and re-thinking optimal job design becomes increasingly important.

References

- Akerlof, George A, "A theory of social custom, of which unemployment may be one consequence," *The quarterly journal of economics*, 1980, 94 (4), 749–775.
- Anderson, Ashton, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec, "Steering user behavior with badges," in "Proceedings of the 22nd international conference on World Wide Web" ACM 2013, pp. 95–106.
- Ashraf, Nava, Oriana Bandiera, and Scott S Lee, "Awards unbundled: Evidence from a natural field experiment," *Journal of Economic Behavior & Organization*, 2014, 100, 44–63.
- Bandiera, Oriana, Iwan Barankay, and Imran Rasul, "Team incentives: Evidence from a firm level experiment," *Journal of the European Economic Association*, 2013, 11 (5), 1079–1114.

- Baumeister, RR**, “The self (In DT Gilbert, ST Fiske, & G. Lindzey (Eds.). *The handbook of social psychology* (Vol. 1, pp. 680–740),” *NY: McGraw-Hill*, 1998.
- Bellé, Nicola**, “Performance-related pay and the crowding out of motivation in the public sector: A randomized field experiment,” *Public Administration Review*, 2015, 75 (2), 230–241.
- Benabou, Roland and Jean Tirole**, “Intrinsic and extrinsic motivation,” *The review of economic studies*, 2003, 70 (3), 489–520.
- Bénabou, Roland and Jean Tirole**, “Incentives and prosocial behavior,” *American economic review*, 2006, 96 (5), 1652–1678.
- Bernheim, B Douglas**, “A theory of conformity,” *Journal of political Economy*, 1994, 102 (5), 841–877.
- Besley, Timothy and Maitreesh Ghatak**, “Status incentives,” *American Economic Review*, 2008, 98 (2), 206–11.
- Blader, Steven, Claudine Madras Gartenberg, and Andrea Prat**, “The contingent effect of management practices,” *Columbia Business School Research Paper*, 2016, (15-48).
- Bradler, Christiane, Robert Dur, Susanne Neckermann, and Arjan Non**, “Employee recognition and performance: A field experiment,” *Management Science*, 2016, 62 (11), 3085–3099.
- Buelens, Marc and Herman Van den Broeck**, “An analysis of differences in work motivation between public and private sector organizations,” *Public administration review*, 2007, 67 (1), 65–74.
- Buurman, Margaretha, Josse Delfgaauw, Robert Dur, and Seth Van den Bossche**, “Public sector employees: Risk averse and altruistic?,” *Journal of Economic Behavior & Organization*, 2012, 83 (3), 279–291.
- Cameron, A Colin, Jonah B Gelbach, and Douglas L Miller**, “Bootstrap-based improvements for inference with clustered errors,” *The Review of Economics and Statistics*, 2008, 90 (3), 414–427.
- Cassar, Lea and Stephan Meier**, “Nonmonetary Incentives and the Implications of Work as a Source of Meaning,” *Journal of Economic Perspectives*, 2018, 32 (3), 215–38.
- Chen, Mingliang, Qingguo Ma, Minle Li, Hongxia Lai, Xiaoyi Wang, and Liangchao Shu**, “Cognitive and emotional conflicts of counter-conformity choice in purchasing books online: an event-related potentials study,” *Biological Psychology*, 2010, 85 (3), 437–445.
- Chetty, Raj, John N Friedman, and Jonah E Rockoff**, “Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood,” *American economic review*, 2014, 104 (9), 2633–79.
- Cooley, Charles Horton**, “Looking-glass self,” *The production of reality: Essays and readings on social interaction*, 1902, 6.
- Crewson, Philip E**, “Public-service motivation: Building empirical evidence of incidence and effect,” *Journal of public administration research and theory*, 1997, 7 (4), 499–518.
- Crutzen, Benoît SY, Otto H Swank, and Bauke Visser**, “Confidence management: on interpersonal comparisons in teams,” *Journal of Economics & Management Strategy*, 2013, 22 (4), 744–767.

- Deci, Edward L**, “Effects of externally mediated rewards on intrinsic motivation.,” *Journal of personality and Social Psychology*, 1971, 18 (1), 105.
- Delfgaauw, Josse, Robert Dur, Joeri Sol, and Willem Verbeke**, “Tournament incentives in the field: Gender differences in the workplace,” *Journal of Labor Economics*, 2013, 31 (2), 305–326.
- Dohmen, Thomas and Armin Falk**, “You get what you pay for: Incentives and selection in the education system,” *The Economic Journal*, 2010, 120 (546), F256–F271.
- Dufo, Esther and Rema Hanna**, “Monitoring works: Getting teachers to come to school,” Technical Report, National Bureau of Economic Research 2005.
- Eberts, Randall W, Kevin Hollenbeck, and Joe A Stone**, “Teacher performance incentives and student outcomes,” 2000.
- Fehr, Ernst and Armin Falk**, “Wage rigidity in a competitive incomplete contract market,” *Journal of political Economy*, 1999, 107 (1), 106–134.
- Felson, RB**, “The (somewhat) social self: how others affect self-appraisals’, in (J. Suis, ed.), *Psychological Perspectives on the Self: The Self in Social Perspective*, vol. 4,” 1993.
- Firestone, William A and James R Pennell**, “Teacher commitment, working conditions, and differential incentive policies,” *Review of educational research*, 1993, 63 (4), 489–525.
- Fischer, Paul and Steven Huddart**, “Optimal contracting with endogenous social norms,” *American Economic Review*, 2008, 98 (4), 1459–75.
- Frey, Bruno**, “Not Just for They Money: An Economic Theory of Personal Motivation,” 1997.
- Fryer, Roland G**, “Teacher incentives and student achievement: Evidence from New York City public schools,” *Journal of Labor Economics*, 2013, 31 (2), 373–407.
- Gallus, Jana**, “Fostering public good contributions with symbolic awards: A large-scale natural field experiment at Wikipedia,” *Management Science*, 2016, 63 (12), 3999–4015.
- and **Bruno S Frey**, “Awards as non-monetary incentives,” in “Evidence-based HRM: a Global Forum for Empirical Scholarship,” Vol. 4 Emerald Group Publishing Limited 2016, pp. 81–91.
- Georgellis, Yannis, Elisabetta Iossa, and Vurain Tabvuma**, “Crowding out intrinsic motivation in the public sector,” *Journal of Public Administration Research and Theory*, 2010, 21 (3), 473–493.
- Gibbons, Robert**, “Incentives in organizations,” *Journal of economic perspectives*, 1998, 12 (4), 115–132.
- Glewwe, Paul, Nauman Ilias, and Michael Kremer**, “Teacher incentives,” *American Economic Journal: Applied Economics*, 2010, 2 (3), 205–27.
- Gneezy, Uri and Aldo Rustichini**, “A fine is a price,” *The Journal of Legal Studies*, 2000, 29 (1), 1–17.
- and — , “Pay enough or don’t pay at all,” *The Quarterly Journal of Economics*, 2000, 115 (3), 791–810.
- Goodhart, Charles AE**, “Problems of monetary management: the UK experience,” in “Monetary Theory and Practice,” Springer, 1984, pp. 91–121.

- Goodman, Sarena F and Lesley J Turner**, “The design of teacher incentive pay and educational outcomes: Evidence from the New York City bonus program,” *Journal of Labor Economics*, 2013, 31 (2), 409–420.
- Grant, Adam M and Francesca Gino**, “A little thanks goes a long way: Explaining why gratitude expressions motivate prosocial behavior.,” *Journal of personality and social psychology*, 2010, 98 (6), 946.
- Gubler, Timothy, Ian Larkin, and Lamar Pierce**, “Motivational spillovers from awards: Crowding out in a multitasking environment,” *Organization Science*, 2016, 27 (2), 286–303.
- Hanushek, Eric A et al.**, “Teacher characteristics and gains in student achievement: Estimation using micro data,” *American Economic Review*, 1971, 61 (2), 280–288.
- Harrison, Glenn W and John A List**, “Field experiments,” *Journal of Economic literature*, 2004, 42 (4), 1009–1055.
- Holmstrom, Bengt and Paul Milgrom**, “Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design,” *JL Econ. & Org.*, 1991, 7, 24.
- Hoogveld, Nicky and Nick Zubanov**, “The power of (no) recognition: Experimental evidence from the university classroom,” *Journal of Behavioral and Experimental Economics*, 2017, 67, 75–84.
- Jacob, Brian A and Steven D Levitt**, “Rotten apples: An investigation of the prevalence and predictors of teacher cheating,” *The Quarterly Journal of Economics*, 2003, 118 (3), 843–877.
- Kamphorst, Jurjen JA and Otto H Swank**, “Don’t demotivate, discriminate,” *American Economic Journal: Microeconomics*, 2016, 8 (1), 140–65.
- Kohn, Alfie**, *Punished by Rewards.: The Trouble with Gold Stars, Incentive Plans, A’s, Praise, and Other Bribes*, Houghton Mifflin Harcourt, 1999.
- Kosfeld, Michael and Susanne Neckermann**, “Getting more work for nothing? Symbolic awards and worker performance,” *American Economic Journal: Microeconomics*, 2011, 3 (3), 86–99.
- Lazear, Edward P**, “The power of incentives,” *American Economic Review*, 2000, 90 (2), 410–414.
- Leigh, Andrew**, “Teacher pay and teacher aptitude,” *Economics of education review*, 2012, 31 (3), 41–53.
- Lourenço, Sofia M**, “Monetary incentives, feedback, and recognition—Complements or substitutes? Evidence from a field experiment in a retail services company,” *The Accounting Review*, 2015, 91 (1), 279–297.
- Moldovanu, Benny, Aner Sela, and Xianwen Shi**, “Contests for status,” *Journal of political Economy*, 2007, 115 (2), 338–363.
- Muralidharan, Karthik and Venkatesh Sundararaman**, “Teacher performance pay: Experimental evidence from India,” *Journal of political Economy*, 2011, 119 (1), 39–77.
- Nelson, Bob**, *1501 ways to reward employees*, Workman Publishing, 2012.
- Pop-Eleches, Cristian and Miguel Urquiola**, “Going to a better school: Effects and behavioral responses,” *American Economic Review*, 2013, 103 (4), 1289–1324.
- Rogers, T and E Frey**, “Changing behavior beyond the here and now,” 2016.

Sanders, Michael, Aisling Ni Chonaire et al., ““Powered to Detect Small Effect Sizes”: You keep saying that. I do not think it means what you think it means.” Technical Report, Department of Economics, University of Bristol, UK 2015.

Sliwka, Dirk, “Trust as a signal of a social norm and the hidden costs of incentive schemes,” *American Economic Review*, 2007, 97 (3), 999–1012.

Springer, Matthew G, Dale Ballou, Laura Hamilton, Vi-Nhuan Le, JR Lockwood, Daniel F McCaffrey, Matthew Pepper, and Brian M Stecher, “Teacher Pay for Performance: Experimental Evidence from the Project on Incentives in Teaching (POINT).” *Society for Research on Educational Effectiveness*, 2011.

Stajkovic, Alexander D and Fred Luthans, “Behavioral management and task performance in organizations: conceptual background, meta-analysis, and test of alternative models,” *Personnel Psychology*, 2003, 56 (1), 155–194.

Swank, Otto H and Bauke Visser, “Motivating through delegating tasks or giving attention,” *The Journal of Law, Economics, & Organization*, 2006, 23 (3), 731–742.

Appendix

A.1 Data Collection and Time Line

This experiment follows roughly 900 teachers in 39 Romanian schools over the course of an entire academic year, from September 2017 to August 2018. All the 39 schools in the experiment use an online management platform designed to monitor student performance. Between the 1st of September 2017 and the 21st of January 2018, data on the baseline performance of teachers is collected, and used to compute student performance gains.

Based on their performance, teachers are ranked across all schools, within their own subject. Those teachers who are in the top 25% best performers within their subject are labelled as ‘Top performers’ and qualify for receiving public praise. The 39 schools are then randomly assigned between a ‘Treatment’ group and a ‘Control’ group, stratified on the baseline performance of students (first grade in the beginning of school year) and teachers (pre-intervention performance gains), and on the size of the school. The treatment assignment of schools remains unchanged throughout the entire academic year. There is no selection into the treatment, and no schools opt out of receiving the messages following the first intervention.

The first intervention takes place on the 22nd of January 2018, following the end of the Christmas break. The platform managers post the messages on the platform page of each of the treated schools. The messages posted in each school are identical in terms of content. The only source of variation in the messages is the names of the top performing teachers within each school. The message is only visible to teachers, parents, and students within that school. The message is posted on the main page of the platform, visible immediately after logging-in. An additional email is sent by the platform managers to all teachers in the school, reminding them to read the public message and providing them with a link to the original post.

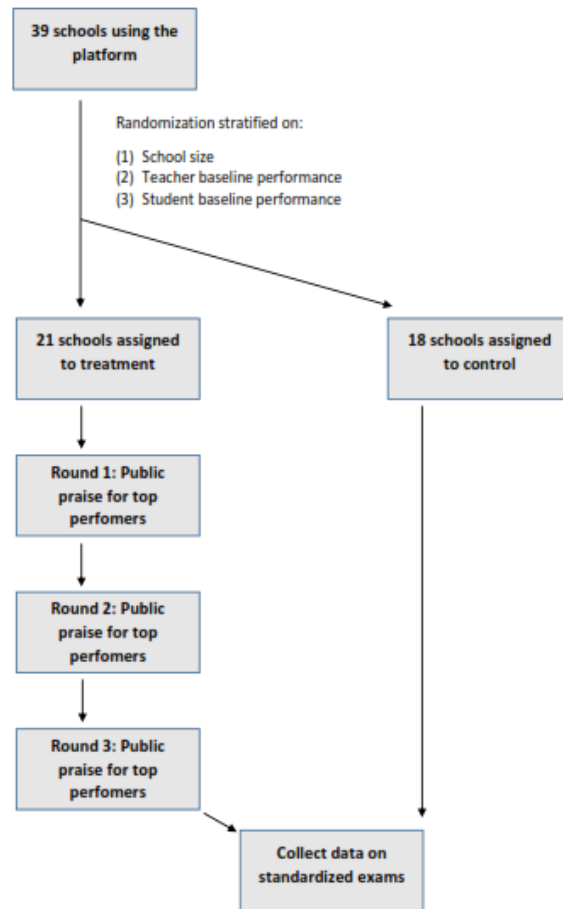
The second intervention takes place on the 20th of March 2018, two months after the first round of messages is sent. Student grades between the 22nd of January 2018 and 20th of March 2018 are used to calculate the new performance gains for all teachers across all schools. Teachers were ranked again based on this new performance gain, and labelled as ‘Top performers’ if they are among the 25% best performing teachers (within their own subject). Top performing teachers in treated schools are publicly praised in a new round of messages posted on the 20th of March 2018.

The third and final intervention takes place on the 15th of May, two months after the second round of messages. Analogous, a final round of public messages is posted in all the treated schools. Finally, teacher performance gains were calculated again between the 15th of May 2018 and the end of the academic year.

In the week of 11th of June 2018, students finishing secondary school (aged 14) undertake high stake standardized exams, which are anonymously graded (‘Examen de capacitate’). In the week of 25th of June 2018, students finishing high school (aged 18) undertake high-stake standardized exams which are anonymously graded (‘Examen de Bacalaureat’).

Figure A.1 presents a schematic overview of the experiment design and timeline.

Figure A.1.: Experiment time-line



A.2 PG robustness checks

Table A.2.1: Relationship between current learning and previous learning

	Pre-treatment PG
Last year's PG	0.079 (0.060) [0.096*]
N	371
F-value	4.39
R-squared	0.19

Notes: The dependent variable is pre-intervention PG expressed in standard deviations. Last year's PG is also expressed in standard deviations. OLS regression controls for student gender composition, average baseline performance of a teacher's students, subject fixed effects, degree of urbanization, being a publicly-funded school, school size, baseline student and teacher quality at the school level, and region fixed effects. In parentheses, heteroskedasticity robust standard errors are presented. Since the analysis is performed on only 20 schools, in brackets the p-value from applying the wild bootstrap procedure on standard errors clustered at the school level are presented (Cameron et al (2008)). Significance levels: *** $p < .01$, ** $p < .05$, * $p < .1$.

Table A.2.2: Pre-treatment PG and learning, for new and old students

	Pre-treatment PG	Pre-treatment PG
Base perf.	-0.626*** (0.027)	-0.630*** (0.027)
New students		0.047 (0.033)
Base perf. * New students		0.033 (0.010)
N	48,547	48,547
F-value	84.64	86.01
R-squared	0.32	0.32

Notes: The dependent variable is pre-intervention PG at the student level, expressed in standard deviations. The baseline performance of students is also expressed in standard deviations. 'New students' is a dummy variable which takes value one for students who just joined the school and 0 otherwise. OLS regressions control for student gender, average baseline student performance, subject fixed effects, degree of urbanization, being publicly-funded school, school size, baseline student and teacher quality at the school level, and region fixed effects. In parentheses, heteroskedasticity robust standard errors are presented, clustered at the school level. Since the same teacher can teach across different classes both new and recurring students, the analysis is performed at the student level. Significance levels: *** $p < .01$, ** $p < .05$, * $p < .1$.

A.3 Full text of intervention

The intervention text, original in Romanian language, is posted by the platform managers on the front page of the website, visible to all teachers, parents and students immediately after logging-in.

Unannounced public praise message:

”Dear Teachers,

We are interested in how the performance of this school’s students is improving over time, since we want to encourage progress in education.

One way to measure the progress of students, is to see how much their grades improved since the beginning of the year. For a number of subjects (Mathematics, Romanian, English, Biology, Chemistry, Physics, History, Geography and Computer Science) we have looked at the improvement in student grades across all the schools that implement the *[platform name]* school management solution.

We are happy to announce that a number of teachers in your school are among the top 25% performers for their subject, across all the schools in our database. For these subjects, their student’s grades have improved the most since the beginning of the semester, as compared to the grades of students from other schools! These teachers are:

[Teacher 1 name]

.....

[Teacher n name]

We would like to thank these teachers in particular for their contribution!

In the future we plan to send such messages more often, to show our gratitude towards your hard work!

Best,”

Announced and repeated public praise message:

”Dear Teachers,

As you know, in the past we have analyzed, for a number of subjects (Mathematics, Romanian, English, Biology, Chemistry, Physics, History, Geography, English and Computer Science), the improvement in student grades across all the schools that implement the *[platform name]* school management solution.

We have now repeated this analysis. We are happy to announce that a number of teachers in your school are among the top 25% performers for their subject, across all the schools in our database. For these subjects, their student’s grades have improved the most over the last 2 months, as compared to the grades of students from other schools! These teachers are:

[Teacher 1 name]

.....

[Teacher n name]

We would like to thank these teachers in particular for their contribution!

In the future we plan to send such messages more often, to show our gratitude towards your hard work!

Regards,”

Additionally, each teacher in a treated school is sent a reminder about the public message, through a personal e-mail from the platform managers. This measure is implemented to ensure that the treatment is as visible as possible.

Following the first intervention, the private message sent to all teachers is:

”Hello,

We are pleased to announce that for a number of subjects we have reviewed the increase in student performance in schools which implement the *[platform name]* school management solution.

Based on this analysis, *[number teachers]* teachers in your school are among the top 25% teachers in existing schools in our database!

If you want to see who these teachers are (or if you are one of them) you can see the list here: *[link to public message]*

Regards,”

Following the second and the third intervention, the private message sent to all teachers is:

”Hello,

We are pleased to announce that we have reviewed again the increase in student performance over the past 2 months. This analysis included all the schools which implement the *[platform name]* school management solution.

Based on this analysis, *[number teachers]* teachers in your school are among the top 25% teachers in existing schools in our database!

If you want to see who these teachers are (or if you are one of them) you can see the list here: *[link to public message]*

Regards,”

A.4 Attrition

Attrition occurs when a teacher did not record any grades, for any of their students, during one experimental period. As such, for this teacher, no performance gain can be calculated. This type of attrition happens if a teacher did not test their students within a given period, or if a teacher stopped using the system all together. From the 855 teachers in the original sample, for 4% ($n = 34$) performance gains cannot be calculated in period 2, 7% ($n = 63$) in period 3, and 3% ($n = 29$) in the last period. For roughly 86% of teachers performance gains can be calculated throughout the intervention, making the said attrition rate reasonably low.

Table A.4 below shows the results from the Hotelling's T-squared test for multivariate data. The test verifies whether two sets of means are equal to each other across two groups, namely between a group of teachers who opt-out by recording no grades ('Attrition') and a group of teachers who do not opt-out ('No attrition'). The test has the advantage of jointly testing multiple variables at the same time, in this case the treatment status, being a top performer, and the interaction between the two. In the case of only one variable, the test reduces to a standard t-test. According to the results in Table A4, attrition each round does not depend on either treatment status, on being a top performer, or on the interaction between the two.

Table A.4: Balance test for joint mean differences between the 'Attrition' and 'No attrition' group, each round

Round 1	No attrition	Attrition
Treatment	0.55	0.44
Top performer	0.25	0.32
Treatment * Top performer	0.14	0.15
N	821	34
F-value joint difference	1.15	
P-value joint difference	0.33	
Round 2	No attrition	Attrition
Treatment	0.55	0.54
Top performer	0.26	0.21
Treatment * Top performer	0.12	0.14
N	758	63
F-value joint difference	1.53	
P-value joint difference	0.21	
Round 3	No attrition	Attrition
Treatment	0.54	0.76
Top performer	0.24	0.17
Treatment * Top performer	0.12	0.14
N	729	29
F-value joint difference	1.93	
P-value joint difference	0.12	

Notes: Columns show mean differences between the 'No attrition' and the 'Attrition' groups. Significance levels: *** $p < .01$, ** $p < .05$, * $p < .1$.

A.5 Unannounced public praise and attendance

In this section I investigate the effects of unannounced praise on school attendance. Teachers are not praised on the basis of class attendance. However, as teacher performance changes as a result of unannounced praise, students can also become more or less likely to attend class. Furthermore, when the platform managers send a public praise message, they also signal that they are paying attention. This might induce teachers to feel monitored, and take active steps to increase the attendance rate. Table A.5.1 presents the cumulative number of skipped classes among all the students of a teacher, up to that period in time.

Table A.5.1: Cumulative number of skipped classes among all the students of a teacher

	Treatment	Control	N
Pre-treatment	0.754	0.788	855
Post unannounced praise	1.229	1.064	821
Post announced praise 1	3.472	3.577	756
Post announced praise 2	4.136	4.556	729

Notes: Columns show mean differences in cumulative number of skipped classes between the 'Treatment' and the 'Control' groups. The cumulative number of skipped classes is calculated at the teacher level, across all of their students.

Pre-intervention attendance does not differ across treatment and control groups. After two rounds of the intervention, the students of teachers in the treatment group appear to have fewer classes skipped on average. However, both the level and the variation in skipped classes is very low, suggesting that teachers have very limited control over the measure, and little space for improvement. On average, by the end of the year a teacher will record less than five skipped classes across all of her students. Given the limited variation in attendance, the minimally detectable effects are also large.

I estimate equation (1) where the main outcome variable is average skipped classes per teacher, as accumulated up to time t . Table A.5.2 shows the average treatment effects on attendance, following unannounced praise. The point estimates indicate that the treatment does not have a significant impact on class attendance, and the coefficients are small and imprecisely estimated.

Table A.5.2: The effect of unannounced public praise on attendance

Skipped classes	Skipped classes	Skipped classes
(α_1) Treatment	0.185 (0.290)	0.247 (0.259)
(α_2) Top performer		0.213 (0.213)
(α_3) Treatment * Top performer		-0.243 (0.255)
Teacher Fixed Effects	yes	yes
Time Fixed Effects	yes	yes
N	821	821
F-value	3.45	1.96

Notes: The dependent variable is the cumulative number of skipped classes at the teacher level, two months after the first intervention. In parentheses, heteroskedasticity robust standard errors are presented, clustered at the school level. Significance levels: *** $p < .01$, ** $p < .05$, * $p < .1$.

A.6 Heterogeneous treatment effects for bottom performers

I estimate the following equation:

$$Perf_{i,t=1} = \alpha_0 + \alpha_1 T_i + \alpha_2 Quant_{i,t} + \alpha_3 T_i * Quant_{i,t} + \mu_i + \tau_t + \varepsilon_{i,t} \quad (6)$$

where $Perf_{i,t+1}$ is the new performance gain for teachers following the first intervention, and $Quant_{i,t}$ is a set of dummies for each of the four quantiles of the teachers's performance distribution. Table A.6 below shows the results from estimating equation (6) for teachers in the 1st quantile who qualified for praise (the top 25%), teachers in the 2nd quantile (between 25% and 50% of the performance distribution), teachers in the 3rd quantile (between 50% and 75% of the performance distribution), and teachers in the 4th quantile (the bottom 75% of the distribution) as compared to similar teachers in the control group.

Table A.6 shows that teachers in treated schools at different quantiles of the bottom 75% performance distribution do not respond differently following the intervention. A t-test of joint equality of the coefficients on the three bottom quantiles returns an F-value of 1.90 and a p-value of 0.16.

Table A.6: The effect of unannounced public praise on PG, for different quantiles of the performance distribution

	New PG
Treatment	-0.374** (0.160)
Top performers (Q1)	-2.765*** (0.187)
Quantile 2	-1.505*** (0.134)
Quantile 3	-0.928*** (0.176)
Treatment * Top performers (Q1)	0.599** (0.251)
Treatment * Quantile 2	0.199 (0.161)
Treatment * Quantile 3	0.154 (0.220)
Teacher Fixed Effects	yes
Time Fixed Effects	yes
N	821
F-value	152.71

Notes: The dependent variable is the PG calculated two months after the first intervention, expressed in standard deviations. In parentheses, heteroskedasticity robust standard errors are presented, clustered at the school level. Significance levels: *** $p < .01$, ** $p < .05$, * $p < .1$.

A.7 Balance tests for final years students

Table A.7 reports the coefficients, standard errors and p-values from regressing a number of controls on a dummy variable which takes value one if a student is in the final year, and zero otherwise. With one exception, final year students do not appear to be different across any dimension, neither are they more likely to be over-sampled from the treated group. Final year students are slightly less likely to be sampled from schools that have some private funding. However, this is mechanically determined by the fact that these schools are mostly focused on secondary education, and typically do not offer classes for final year high school students. The main specifications control for school funding.

Table A.7: Differences between final year students and non-final year students

Variable	Is a final year student		
	β	σ	p-value
Baseline performance	0.011	0.008	0.172
Pre-treatment PG	0.007	0.006	0.264
Female student	-0.011	0.006	0.352
In treated group	0.009	0.026	0.745
Urban school	0.030	0.031	0.336
Private funding	-0.089***	0.030	0.005
Randomization variables			
School size	-0.001	0.001	0.493
Baseline teacher performance	-0.091	0.075	0.234
Baseline student performance	0.013	0.016	0.392
F- value	2.49		
P-value	0.024		
R-squared	0.004		
N	48,101		

*Notes: The dependent variable is an indicator taking value 1 if a student is in their final year (aged 14 and 18) and 0 otherwise. Student baseline performance and pre-treatment PG are expressed in standard deviations. In the first column, coefficients from an OLS regression of the dependent variable on controls (β) are presented. In the second column, heteroskedasticity robust standard errors (σ) are presented, clustered at the school level. In the final column, the associated p-value on each coefficient is displayed. Significance levels: *** $p < .01$, ** $p < .05$, * $p < .1$.*

A.8 Announced public praise and attendance

Table A.8 shows the per-period treatment effect on student attendance throughout the remainder of the school year. The intervention has no significant effect on attendance.

Table A.8: The effect of unannounced and announced public praise on attendance

	Skipped Classes	SE
Treatment		
($\gamma_{1,1}$) Round 1	0.203	(0.300)
($\gamma_{1,2}$) Round 2	-0.130	(0.547)
($\gamma_{1,3}$) Round 3	-0.162	(0.657)
Recognition		
($\gamma_{2,1}$) Round 1	0.287	(0.125)
($\gamma_{2,2}$) Round 2	-0.112	(0.125)
($\gamma_{2,3}$) Round 3	0.380	(0.387)
Treatment * Recognition		
($\gamma_{3,1}$) Round 1	-0.058	(0.246)
($\gamma_{3,2}$) Round 2	0.316	(0.259)
($\gamma_{3,3}$) Round 3	-0.783	(0.503)
Teacher Fixed Effects	yes	
Time Fixed Effects	yes	
N	821	
F-value	20.93	

*Notes: The dependent variable is the cumulative number of skipped classes at the teacher level, two months after each intervention. In parentheses, heteroskedasticity robust standard errors are presented, clustered at the school level. Significance levels: *** $p < .01$, ** $p < .05$, * $p < .1$.*

However, by the third round, praised teachers appear to record less skipped classes for their students. While the coefficient is not significant, it is quite sizable. On the other hand, no consistent pattern is observed in the control group, in line with the fact that there is little variation in skipped classes, and with the fact that most students appear to attend regularly regardless of teacher quality.

A.9 Unannounced public praise and exam choice

To test whether students are more likely to select an elective exam if their teacher was praised in the first (unannounced) intervention, I define subject choice as a dummy variable which takes value 0 if the subject is part of the track specific electives, but the student does not choose to undertake the exam. Alternatively, the variable takes value 1 if the subject is part of the track specific electives and the student chooses to undertake the exam.

As the regression only includes schools who offer a high-school program, the number of clusters drops to 22. To ensure that the standard errors are not biased by the small number of clusters I bootstrap standard errors using the wild bootstrap procedure according to Cameron, Gelbach, and Miller (2008), with 5000 replications. Table A.9 below also reports the coefficients and associated the p-values from this exercise, confirming that the treatment does not influence elective choice.

Table A.9: Unannounced public praise and exam choice

	Subject chosen
Treatment	0.043 (0.045) [0.395]
Top performer	0.053 (0.109) [0.647]
Treatment * Top performer	-0.096 (0.126) [0.496]
N	4,563
F-value	9.16
R-squared	0.027

*Notes: The dependent variable is an indicator which takes value 1 if the student chose the respective elective course as a final exam, and 0 otherwise. OLS regression controls for baseline student performance, student gender, class type fixed effects, degree of urbanization, being a publicly-funded school, school size, baseline student and teacher quality at the school level, past year exam performance at the school level, and region fixed effects. In parentheses, heteroskedasticity robust standard errors are presented, clustered at the school level. Significance levels: *** $p < .01$, ** $p < .05$, * $p < .1$.*