A market for integrity
The use of competition to reduce bribery in education

Ivan Soraperra*[1], Nils Köbis†[1], Shaul Shalvi[2], Sonja Vogt[3], Charles Efferson‡[3], and Theo Offerman[2]

[1]Center for Humans and Machines, Max Planck Institute for Human Development, Germany
[2]FEB - Amsterdam School of Economics, University of Amsterdam, The Netherlands
[3]Faculty of Business and Economics, University of Lausanne, Switzerland

Abstract

Bribery to attain academic credentials is a widespread problem around the globe. It reinforces inequality and lays the foundation for a norm of undermining skill, achievement, and productivity. One recommended solution is to raise teacher salaries. Even without monitoring and punishment, this could ameliorate the situation by moving teachers away from the poverty threshold and shifting the balance of decision-making forces in favour of intrinsically prosocial motives. As an alternative solution, we suggest a piece-rate scheme that rewards teachers for the students they attract to their schools. Using a game theoretic model and a pre-registered experiment, we compared the two mechanisms. Our results show that a salary increase has little or no impact on bribery, but the piece rate substantially reduces it. The piece rate does this by creating a market that transfers power from teachers to students and thus recruits endogenous forces to reduce bribe extraction by teachers. These findings provide initial evidence for the value of correctly incentivising integrity. Interventions inspired by these findings would be best implemented in cases where the government cannot reliably monitor and enforce anti-corruption measures.

Keywords: Education, bribing, corruption, experiment, market design
JEL Classification: D73, D82, C90
Acknowledgement: This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement: ERC-StG-637915).
Competing interests: The author(s) declare none.

*Corresponding author. Shared first authorship. Email: soraperra@mpib-berlin.mpg.de
†Shared first authorship. Email: n.c.kobis@gmail.com
‡Shared senior authorship.

Introduction

Corruption in education brings considerable social costs (Hallak & Poisson 2007). Widespread practices like students bribing teachers undercut equal opportunities for knowledge acquisition (Rothstein 2011), and this, in turn, have negative downstream consequences that are diverse and far-reaching (Conti et al. 2010). Bribery in the education system reinforces socioeconomic inequalities (Hallak & Poisson 2007), it corrupts people's concept of fairness (Heyneman et al. 2008), and it contributes to the emigration of talent (Cooray & Schneider 2016).

While difficult to document, attaining credentials with bribes instead of effort is a widespread problem that represents the norm in many countries. Recent studies provide evidence suggesting that student bribing examiners is a common practice in Eastern European countries. The introduction of a nationwide campaign to curb bribing at high-school final exams in Romania and Moldova, for instance, caused a staggering drop of about 40 percentage points in the passing rates, suggesting that bribing was indeed a ubiquitous phenomenon (Borcan et al. 2014, 2017). Furthermore, evidence from the Caucasus shows that academic fraud, including paying bribes and sending gifts to teachers, is not only commonly admitted but has significant consequences for the labor market (Mavisakalyan & Meinecke 2016). Finally, survey data from the Global Corruption Barometer conducted by the world's leading anti-corruption organization Transparency International, shows that many students reported paying bribes to obtain services in developing countries as well (Transparency International 2013). In Liberia, for example, 75% of respondents openly admitted they had to pay bribes in the preceding 12 months. Similar answers were given in India and Sierra Leone with 48% and 62% of the respondents making the same admission, respectively. Because of the large societal costs of this kind of corruption, donor organizations, and government bodies have invested vast sums of money in improving education worldwide, with the reduction of corruption a critical and widely acknowledged prerequisite (Munro & Kirya 2020, Global Partnership for Education 2020).

This paper introduces a theoretical framework that models bribery transactions between students (i.e., families) and schools. We study bribery in the absence of reliable punishment institutions (Dorrough et al. 2023)—reflecting the common situation of impunity present in many developing countries where punishment mechanisms are often not feasible because the resources to monitor a large and dispersed population of public officials are simply not available (Stephenson 2020, Fisman & Golden 2017). In our framework, when teachers solicit bribes, students face a social dilemma (Rothstein 2011, Köbis et al. 2016). They can pay to receive a good grade or diploma with little effort, but bribes reduce the perceived value of a diploma in the labor market. For most students, bribes are individually beneficial but socially costly. When teachers earn meager salaries and solicit bribes this can lead to a bad equilibrium in that bribing becomes the widespread norm. In this setup, we test the mechanisms behind two interventions to reduce high levels of corruption. Theoretically and in an experiment, we compare the commonly proposed policy of a fixed-wage increase for teachers (Rose-Ackerman & Palifka 2016, Tanzi 1998, Fisman & Golden 2017) to a new alternative of a piece-rate scheme to fight bribery.

The mechanism behind the fixed-wage increase is that teachers accept bribes to increase their insufficient salaries to support their families. Grounded in Akerlof & Yellen's (1990) fair-wage hypothesis, the thinking is that teachers can only afford to behave morally when their base wage is set at an adequate level (see also Becker & Stigler 1974). As an alternative, the piece-rate scheme pays teachers according to the number of students that they attract (Friedman 1955, 1962). The mechanism behind this approach to fighting bribery is that schools will be forced to curb bribery levels when competing for students. If many students acquire a diploma through bribery instead of hard work, the value of a diploma will be eroded. This makes schools whose diplomas have lost most of their reputational value unattractive to the students which, by voting with their feet, can induce teachers to stop soliciting bribes. Hence, it creates an endogenous punishment mechanism (Sutter et al. 2010) that can help to reduce corruption when no reliable punishment institutions are in place, as is often the case when corruption widely occurs

(Voors et al. 2018)

The available evidence for both interventions is inconclusive: whereas our literature review shows that the empirical evidence on the effectiveness of the fixed-wage increase is mixed, empirical investigation into the success of piece-rate schemes to fight bribery and corruption is lacking altogether. So far, no research has assessed the effectiveness and mechanisms of the two schemes when introduced under identical circumstances. To fill this gap, we conducted a large pre-registered lab experiment.[1] A major advantage of lab experiments is that they allow to directly measure otherwise unobservable behavior, such as bribery, in a controlled environment (Falk & Heckman 2009) and allow to test the causal links between market competition and bribery that are often endogenous (Bennett et al. 2013).

In our experiment, we first pay all teachers a low fixed-wage to stimulate bribery. After high levels of bribery stabilize and thus a norm of corruption is established, we introduce either a fixed-wage increase or a piece-rate salary scheme. While bribery levels remain high after the introduction of a substantial fixed-wage increase, the piece-rate salary scheme substantially reduces bribery—creating a market for integrity.[2]

The remainder of the paper is structured as follows. In the next section, we review the empirical literature on whether bribery and, more generally, corruption can be reduced by higher wages. The following sections introduce our theoretical model and outline the design and the predictions. Finally, we present the results. Finally, we discuss limitations, policy implications, and challenges for implementation before concluding.

[1]For pre-registration, data, R-scripts, and materials, see Open Science Framework.

[2]It is important to emphasize that we focus on settings in which anti-corruption interventions are most needed, yet at the same most difficult to implement: namely environments characterized by impunity. In many developing countries, the likelihood of getting caught and punished for bribing is very low, either because people do not report corrupt behavior or because when corrupt behavior is reported, there is no prosecution of the persons accused of corruption (Di Tella & Schargrodsky 2003, Lambsdorff & Fink 2006). The proposed piece-rate scheme thus presents an alternative to the frequent attempts to introduce effective punishment institutions to societies with high levels of corruption, which have mostly failed (Mungiu-Pippidi 2017, Fisman & Golden 2017, Rothstein 2011).

Literature review on the use of salaries to fight corruption

To reduce bribery and corruption in education, the literature is unanimous in attributing high importance to adequate salaries for teaching staff (Dolan et al. 2012, Wood & Antonowicz 2011, Chaudhury et al. 2006). Seminal work argues that an unconditional salary increase for public officials leads to more productivity, less shirking, and as a specific case, less corruption (Shapiro & Stiglitz 1984). According to Akerlof & Yellen's (1990) fair-wage hypothesis, people who perceive to get poorly paid might "take what they deserve". Getting a low salary provides a moral justification to balance the scale through other means like extracting bribes.

The literature on using unconditional wage increases to reduce corruption reveals a diverse set of patterns. Increasing public salaries unconditionally decreased corruption in some studies, and it increased corruption in other studies. For example, one correlational study by Van Rijckeghem & Weder (2001) compares the relationship between public wages and perceived levels of corruption for 31 developing countries and low-income OECD countries. The results indicate a weak but robust negative relationship between public sector wages and perceived levels of corruption, even when controlling for national-level GDP, indicators of the "rule of law" and "quality of government". Other cross-national studies on the subject by Rauch & Evans (2000) and Treisman (2000) find no robust correlation between salary increases and corruption. More recently, An & Kweon (2017) used cross-country data of a period of 10 years and found a weak negative relationship between public sector wages and perceived corruption. These studies shed no light on the causality between public wages and corruption.

A closer look at the papers reveals some prominent cases supporting the assumed link between higher public wages and a reduction in corruption. A well-known case is the transformations of Singapore, which saw public salary increase schemes as well as a reformation of the policing institutions that subsequently led to an unprecedented drop in corruption levels (Svensson 2005). In the same spirit, Borcan et al. (2014) find that a wage *cut* leads to *more* bribery. Unexpected policy reforms reduced the salaries of Romanian public teachers by 25%, yet left private teachers' salaries unaffected. According to the authors, teachers in public schools extracted extra income by selling the exam results to students to compensate for the wage loss—a practice that did not occur in the private schools where the teachers' salaries remained constant.

In other cases, public officials seem to accept a wage increase as manna from heaven and refrain from changing their behavior. Most relevant for our study is a recent field experiment in Indonesia by de Ree et al. (2017). They report results from a large-scale randomized experiment in which teachers' wage increase was accelerated in the treated schools. They found a null effect of the large pay increase on student learning outcomes, even though teachers' satisfaction with their income was enhanced while the incidence of teachers holding outside jobs was reduced.

Finally, there is also evidence that an increase in the wages of public officials may worsen the problem. Foltz & Opoku-Agyemang (2015) investigated the impact of a Ghanaian wage reform that saw the police force's salary doubled in 2010. They found that the salary increase led to higher levels of effort to collect bribes by the police officers, who operated more checkpoints. While the total *number* of bribes paid remained the same, the salary increase led to higher bribes being requested and an increase in the total *value* of bribes paid.

The mixed evidence in field data is mirrored in controlled experiments on the fair-wage hypothesis. While some studies reveal no corruption-reducing effects of higher wages (Abbink 2002), others show (small) reductions in bribery (Jacquemet 2005, Barr et al. 2009)—in particular when wage increases are combined with effective punishment (Van Veldhuizen 2013). Of particular importance for our study are the experimental papers by Armantier & Boly (2011, 2013), who combined lab and field experiments to study bribery in the education system in Burkina Faso. In Armantier & Boly (2011), participants were recruited for a part-time job in which they had to grade a set of 20 exam papers. One of the 20 exams came with money and a message stating: "Please, find few mistakes in my exam paper". Among others, they compare a high wage treatment (7000CFA = €10.67) to a baseline wage (5000CFA = €7.62). With a high wage, fewer graders accept the bribe, but there is overall more corrupted grading, meaning that the graders accepting the bribe comply more frequently with the bribers' request to find fewer mistakes. In a later version with the same experimental set-up, Armantier & Boly (2013) compare the strength of the effects across different settings: a laboratory in a developed country (Canada), a laboratory, and a field in a developing country (Burkina Faso). They again find the twofold effect of higher wages: less bribe acceptance but more reciprocation. The direction and magnitude of the effects were statistically indistinguishable across the three settings.

Taken together, the literature on the link between wages and corruption provides only weak support for the hypothesis that higher salaries can reduce bribery. If anything, the negative link specifically occurs when realistic threats of punishment exist. Furthermore, empirical evidence for the effectiveness of using piece-rate payment to reduce bribery in the education sector is lacking altogether.

Recent conceptual work highlights the importance of recognizing how anti-corruption interventions affect existing power structures (Köbis et al. 2022). Indeed, unconditional salary increases may not reliably reduce bribery because they leave existing power relations intact. Those who extract bribes before receiving a raise may have little reason to stop after receiving a raise, especially if a strong norm of corruption is already in place. A possible strategy to disrupt such a situation would be to transfer power to the victims of corruption. With respect to education, the victims in question are the students and parents who favour a quality education unmarred by bribery. Accordingly, echoing recent calls for theory-driven social science (Muthukrishna & Henrich 2019), we developed a formal model that examines how incentives and power structures shape bribery in educational systems without a strong punishment threat.

Theoretical model

In this Section, we present a model of bribery in the education sector and solve it for the case of selfish preferences. In the section "Hypotheses and predictions", we derive behavioral predictions that are specific to the experimental parameters considering both selfish and other-regarding preferences.[3]

To model bribery in education, we analyze interactions between two teachers, T_1 and T_2, and N heterogeneous students. We use the term "teacher" for convenience, but one can just as well interpret the model as representing two schools where teachers within the school make coordinated decisions. In both cases, the general framework takes the following form. Teachers decide whether to solicit bribes from students. Students observe teacher choices and choose whether to go to school or go for their best outside option. The outside option reflects not going to school and instead pursuing other activities or opting for a school with lower passing criteria (which is not present in our model). Conditional on teacher choices, students who choose to go to school are either randomly assigned to teachers, or they can choose their teachers. If a student chooses or is randomly assigned to a teacher who solicits bribes, the student can choose to pay the bribe or not. Events unfold in the following order.

- The N students privately learn their respective individual abilities. A student's ability specifically takes the form of an effort cost, c_j, where $j \in \{1, 2, \ldots, N\}$ indexes students. Effort costs are independent draws from a distribution, $\mathcal{F}$, with support $[0, \bar{c}]$.

- Teachers simultaneously decide to solicit bribes ($B_i = 1$) or not ($B_i = 0$), where $i \in \{1, 2\}$ indexes teacher.

- Students observe teacher decisions and simultaneously make their choices about school. Specifically, if one teacher solicits a bribe, and the other does not, each student chooses either (i) the teacher soliciting bribes, (ii) the teacher not soliciting bribes, or (iii) to bypass school and go for the outside option. If, instead, both teachers make the same choice, either soliciting bribes or not, each student chooses either (a) to go to school or (b) to go for the outside option. In this case, the students choosing to study are randomly split between the two teachers.[4]

[3] In Appendix A, we propose some extensions showing how the predictions vary when some of the assumptions are modified.

[4] Letting n designate the number of students who choose to go to school when both teachers make the same choice, $\lfloor \frac{n}{2} \rfloor$ students are randomly assigned to one teacher and $\lceil \frac{n}{2} \rceil$ to the other teacher.

- Students learn the number of students with each of the two teachers. For students with a teacher soliciting bribes, each student simultaneously decides to pay the bribe ($P_j = 1$) or to refuse ($P_j = 0$). Students with a teacher who is not soliciting bribes do not pay bribes ($P_j = 0$).

Payoffs for the students

If a student chooses to go to school and study hard, the student pays her realized effort cost, c_j, to obtain a diploma worth at most 1 unit. However, if a student chooses a teacher soliciting bribes, the student can pay a bribe, $b > 0$, to obtain a diploma without paying the effort cost. Crucially, a student who bribes imposes a negative externality on other students with the same teacher. She does so because bribes reduce the reputational value of diplomas awarded by the teacher in question. If k students in a class bribe a teacher with n students, the teacher's diplomas each have a value of $1 - \frac{k}{n}a$, where $\frac{a}{n}$ is the reduction in diploma value per student bribing in a class with n students.

The payoff for a student, $\pi_j(P_j, n_b, n, c_j)$, depends on whether the student bribes, P_j, on the number of students in the class, n, on the number of *other* students bribing the teacher, n_b, and on the student's ability, c_j. The payoff for a student who does not bribe is

$$\pi_j(0, n_b, n, c_j) = 1 - c_j - \frac{n_b}{n}a.$$

A student who bribes earns

$$\pi_j(1, n_b, n, c_j) = 1 - b - \frac{1}{n}a - \frac{n_b}{n}a.$$

Given a teacher who does not solicit bribes, no one pays a bribe, and a student's payoff is $\pi_j(0, 0, n, c_j) = 1 - c_j$. Lastly, students can choose to bypass school and go for the outside option with a payoff of π_{out}.

Notice that the reputation cost produced by a student who decides to pay the bribe, i.e., $\frac{a}{n}$, is decreasing with the size of the school. Hence, the larger the school, the less a bribing individual student is affected by a loss of reputation.

The reputation cost reflects the social-dilemma aspect of bribing. All students experience a reputation costs from a student's bribing activity. All students are better off when others do not bribe, but some students will not want to resist the temptation when the possibility is offered.

The model captures two important features of education in many countries. First, markets account for diplomas that have a low value due to bribery and scandal (Luca et al. 2017). In particular, employers can and do discriminate among students based on the reputations of the schools students attend (Heyneman et al. 2008). This implies that a diploma from a corrupt school imposes a cost on the students who attend the school but choose to study hard instead of paying bribes. Such students are prepared to learn, but they must absorb the negative externalities produced by others. Second, students and their families have a good sense of which schools are corrupt and which schools are not, and students can thus decide where to go to school based on this information (Chumacero et al. 2011, Luca et al. 2017).

Payoffs for the teachers

Teacher payoffs depend on the institutional setting. Under the *fixed-wage* scheme, a teacher receives a fixed payment of F. Under the *piece-rate* scheme, a teacher receives a piece-rate payoff, s, for each student who chooses the teacher's class. Regardless of the setting, a teacher collects any bribes students pay in the teacher's class. In general, a teacher's payoff depends on the number of students in the class, n, and the number of students paying bribes in the class, k_b. Under the fixed wage, a teacher's payoff is

$$\pi_i^f(n, k_b) = F + bk_b,$$

and under the piece rate it is

$$\pi_i^{pr}(n, k_b) = sn + bk_b.$$

When a teacher does not solicit bribes, $k_b = 0$ and the payoffs π_i^f and π_i^{pr} change accordingly.

Equilibrium analysis

We start the analysis for the case that we consider most relevant and that forms the basis of the experiment. Appendix A considers various extensions of the model. In the analysis we make the following assumptions and, at the end of this Section, we briefly discuss how the analysis changes if these are not fulfilled.

- Assumption 1: $\pi_{out} < 1 - b - a$. This implies that paying the bribe to obtain a diploma with any reputation is better than skipping school altogether. Hence, students prefer obtaining an educational degree, even a bad one, "for free" rather than having no degree at all.

- Assumption 2: $\bar{c} > 1 - \pi_{out}$. This implies that there are some students that would prefer not to obtain the diploma when they have to work hard to get it.

We focus on symmetric perfect Bayesian Nash equilibria in which all the students use the same strategy when deciding to pay or not to pay a bribe. Given assumptions 1 and 2, the following proposition highlights the conditions to have a unique symmetric perfect Bayesian equilibrium where teachers do not solicit bribes and describes the strategies of the players. The proof of the proposition is reported in Appendix A.1

Proposition. *In the symmetric perfect Bayesian equilibrium, students' strategies are characterized as follows:*

(i_a) when both teachers do not solicit bribes, students with $c_j > 1 - \pi_{out}$ do not go to school and students with $c_j \leq 1 - \pi_{out}$ go to school and are randomly split between the teachers;

(i_b) when one teacher solicits bribes and the other teacher does not, the students with $c_j > b + a$ choose the teacher who solicits bribes and the other students choose the teacher who does not solicit bribes;

(i_c) when both teachers solicit bribes, all the students go to school and they are randomly split between the teachers;

(ii) when $n > 0$ students have chosen a teacher who solicits bribes, the students with $c_j > b + \frac{1}{n}a$ pay the bribe, while the others do not pay the bribe.

In the symmetric perfect Bayesian equilibrium, teachers' strategies depend on the institutional regime and are characterized as follows:

(iii_a) in the fixed-wage *regime teachers solicit bribes;*

(iii_b) in the piece-rate *regime teachers do not solicit bribes provided that:*

- *the loss in reputation is not negligible compared to the distribution of effort costs*

$$\mathcal{F}(a+b) > 1 - \frac{\mathcal{F}(1-\pi_{out})}{2}$$

- *the piece-rate is large enough compared to the level of the bribe $s > t \cdot b$ where*

$$t = \max\left(\frac{E_b}{N\left(\mathcal{F}(a+b) - \frac{1}{2}\right)}, \frac{2\left(1-\mathcal{F}(a+b)\right)}{\mathcal{F}(1-\pi_{out}) + 2\mathcal{F}(a+b) - 2}\right).$$

The intuition for why under the *fixed-wage* regime, bribes are solicited in equilibrium is straightforward: Since teachers are paid a fixed-wage F, soliciting is at least as good as not soliciting bribes. Moreover, if they solicit bribes they have a positive probability of attracting some "bad" students, i.e., those students with $c_j > a + b$, because these students prefer to buy a diploma over studying hard to get a diploma. This and the assumption that a diploma with a bad reputation is better than the outside option imply that soliciting bribes is the dominant option for teachers.

As for the *piece-rate* regime, teachers are paid according to the number of students in their class and, if the piece-rate s is large enough, they find themselves in a prisoner's dilemma where not soliciting bribes is the dominant strategy. This happens because everyone knows that the bad students, i.e., the ones with $c_j > a+b$, will always choose a teacher soliciting bribes and will pay the bribe. Therefore, to avoid the reputation cost imposed by the bad students, the other students prefer a teacher not soliciting bribes, if available. When the reputation damage of bribes is not too low, there is enough demand for bribe-free schools to guarantee that a teacher who stops soliciting bribes can attract the majority of the students and hence obtain better pay.

In the main analysis, we focus on the simple case where the decision to bribe is binary. The important results carry over to the case where the bribe can be set to any level. In Appendix A.2, we analyze the case of teachers endogenously setting the level of the bribe and identify sufficient conditions to guarantee the existence of an equilibrium where bribes are not solicited in the piece-rate regime. There are two important results worth highlighting when allowing for endogenous bribes. First, under piece-rate one may think that competition would push teachers to giving away diplomas for free (i.e., setting $b = 0$) to attract more students and to cash-in the piece-rate. This is not the case. Indeed, under the conditions discussed in our proposition, teachers have a profitable deviation. By stopping soliciting bribes they can attract more students than by giving away the diploma for free. Second, contrary to the intuition that, with fixed wages, competition and endogenous bribes would be sufficient to drive down bribes and, eventually, stop teachers from soliciting (Shleifer 2004), in our case, this would not be an equilibrium. This is because, by deviating to soliciting a positive bribe, a teacher can always attract some bad students who otherwise would not be in school and collect bribes from them.

——————— Figure 1 about here ———————

The proposition points out that the piece-rate and the externality need to be large enough to stop teachers from soliciting bribes. When this is not the case, there are different equilibria depending on the level of s and $\mathcal{F}(a+b)$. Figure 1 provides an example showing how the set of equilibria changes with these parameters (see Appendix A.1 for further details). When either the externality is small compared to the effort cost of the students or s is small, there is only one equilibrium where both teachers solicit bribes. If s is at an intermediate level, there are two cases: (i) both teachers solicit, and both teachers do not solicit bribes are both equilibria and teachers face a coordination game; (ii) teachers specialize with one teacher soliciting and the other not soliciting bribes, and teachers face a "battle of the sexes" type of game.

We conclude with a short discussion of the assumptions. We think that it is reasonable to assume that some students cannot obtain a diploma when they have to exert effort (Assumption 2). Here, we explain what happens when Assumption 1 does not hold, and a diploma with a bad reputation is worse than the outside option. Without Assumption 1, students would prefer to drop out of school when there are too many bribes paid, and our equilibrium analysis would change only for the case where bribes are solicited. Specifically, it would make bribe solicitation less appealing for the teachers who would lose some students, and it would strengthen the effect of the piece-rate regime.

Experimental design, procedure and hypotheses

Design

The experiment had two treatments: Fixed Wage (FW) and Piece Rate (PR). In each treatment, participants were randomly matched into 24 groups of 10 participants, with two participants playing in the role of the teacher and eight playing in the role of student.

In the experiment, we explicitly labeled participants' roles as teachers and students. In addition, we referred to the bribe as a "motivation fee". The decision to use framed instructions was motivated by the need to facilitate comprehension of the game and to introduce an element of moral cost when soliciting and paying the bribe. Moreover, it is in line with many of the experimental literature on corruption games (see, e.g., Abbink & Hennig-Schmidt 2006, Armantier & Boly 2013, Barr & Serra 2009).

Regardless of the treatment (FW or PR), an experimental session consisted of two parts of 15 periods each. In the first part, participants repeatedly played the stage game explained above under a low fixed wage regime to create clear incentives for teachers to solicit bribes. In this first part, we sought to establish a widespread norm of bribing to create a more realistic setting to test the different interventions in. In period 16, the first period of the second part, we introduced the treatment interventions designed to reduce bribery. Depending on the treatment, the intervention was either a new fixed wage, much higher than before, or a piece rate.

Participants' group membership and role assignment were determined at the beginning of the experiment and were kept unchanged for the entire duration of the experiment. The students' effort cost (see next paragraph), however, was randomly drawn each period. This means that participants in the role of the students could have different effort costs in different rounds of the experiment. When teachers offered the same option, students had to decide whether to go to school—and be randomly assigned to one of the teachers—or not; when teachers offered different options, students had to choose between "the teacher asking for a motivation fee" and "the teacher not asking for a motivation fee".

In detail, teachers received a *low* fixed wage of 40 points for each of the first 15 periods of a session in both treatments. A diploma was worth a maximum of 250 points to students, but a diploma could lose up to 100 points in value when everyone in the class paid a bribe ($a = 100$). Paying a bribe amounted to the student transferring 10 points to the teacher ($b = 10$). Students were one of three types. Good students had a low effort cost of $c_j = 10$ points, intermediate students a medium effort cost of $c_j = 65$ points, and bad students a high effort cost $c_j = 160$ points. These abilities occurred with probabilities 1/6, 2/3, and 1/6, respectively. The outside option was worth $\pi_{out} = 115$ points. Importantly, the fixed wage for teachers in the first part was lower than any possible payoff for students. Our rationale here was to promote a norm of widespread bribery before the intervention in period 16. Such a strong norm of corruption would provide ample scope for one or both of our interventions to reduce bribery, and it would also represent the main scenario of interest for policymakers who intervene to reduce already high levels of bribery.

At the beginning of the second part of a session, in period 16, we introduced one of our two interventions in a between-subjects design. In the fixed-wage treatment, teachers received an unconditional raise to 240 points per period. In the piece-rate treatment, in addition to the previous fixed payment of 40 points, a teacher received a bonus of 50 points for each student choosing her class ($s = 50$). Crucially, the two interventions were neutral in terms of public expenditure in the precise sense that when all students chose to go to school, both payment schemes required an additional 400 points per period to reduce bribery.

Procedure

The experiment was programmed in z-Tree (Fischbacher 2007) and conducted at the Rosario Experimental and Behavioral Economics Lab (REBEL) in November 2017. The sample consisted of 480 students of the Universidad del Rosario—we ran the experiment in Colombia as a country where corruption is widespread (OECD 2016). The sample's average age was 20.00 ($SD = 2.1$), and 57.7% of the participants were female.

When entering the laboratory, participants were randomly assigned to visually isolated computer terminals and were asked to read and sign a participation consent form.[5] Afterward, the instructions of the first part were distributed and then read aloud and before starting, participants had to answer a series of control questions that tested their comprehension of the rules. At the end of the first part, the instructions for the second part were distributed and read aloud. Participants went through a second series of control questions before starting with the second part of the experiment.

Participants were paid 12000 COP as a show-up fee plus the sum of the payoff obtained in 6 randomly selected periods, 3 from the first part and 3 form the second part. The payoff conversion rate was 30 COP for each point. Each session took approximately 2 hours, and the average earnings were 41000 COP ($\approx$ 12.8 USD), which is higher than the salary a participant can earn in a comparable two hours students' job.

Hypotheses and predictions

In the experiment, we will test the following null hypotheses:

H1: In the first part, teachers solicit bribes (and students pay bribes).

[5] All the materials, consent form, and instructions are reported, translated from Spanish, in Appendix B.

H2: In treatment *PR*, the fraction of bribes solicited (and paid) in the second part does not differ from the fraction of bribes solicited (and paid) in the first part.

H3: In treatment *FW*, the fraction of bribes solicited (and paid) in the second part does not differ from the fraction of bribes solicited (and paid) in the first part.

H4: The fraction of bribes solicited (and paid) in the second part of treatment *PR* does not differ from the fraction of bribes solicited (and paid) in the second part of treatment *FW*.

We now derive predictions for the alternative hypotheses in two cases: the standard case of self-interested teachers and the case of teachers who are driven by other-regarding concerns (the derivations of the predictions for this case are reported in Appendix C).

As for the case of self-interested teachers, note that the experimental setup satisfies the assumption of the model. Therefore, according to the Proposition, the experimental parameters induce the following equilibrium behavior. If both teachers do not solicit bribes, bad students ($c_j = 160$) prefer not to go to school, while intermediate ($c_j = 65$) and good ($c_j = 10$) students prefer to go to school and obtain the diploma. If both teachers solicit bribes, good students are better off by not paying the bribe independently of the size of the class; bad students are best off by going to school and paying the bribe independently of the size of the class; intermediate students do not pay the bribe when there are no other students in the class and pay the bribe in all other cases.

If one teacher solicits bribes and the other teacher does not, bad students will choose the teacher soliciting bribes and pay the bribe, independent of what other students do. Expecting this, good students choose the teacher that does not solicit bribes to avoid the externality. Finally, given that good students are choosing the teacher not soliciting bribes and bad students are choosing the teacher soliciting bribes and are paying the bribe, intermediate student prefers to join the class where bribes are not solicited.

Given that teachers anticipate the behaviors of students as described above, we have the following predictions for their behavior. When teachers are paid a fixed-wage, soliciting bribes is a dominant strategy. When teachers are paid a piece-rate s for each student choosing their class, teachers will refrain from soliciting bribes.

Therefore, in accordance with the equilibrium strategies, we predict that H1 and H3 will not be rejected. Yet, H2 will be rejected for the alternative hypothesis that the fraction of bribes solicited (and paid) is smaller in the second part than in the first part. Also, H4 will be rejected in favor of the alternative hypothesis that in the second part the fraction of bribes in treatment FW exceeds bribe levels in treatment PR.

These predictions show that, after a pay-raise that brings their salary to a fair level, self-interested teachers have no incentive to stop soliciting bribes. This, however, may not necessarily happen if teachers are concerned with the fairness of the payoff distribution.

As for the case of teachers with other-regarding preferences, we provide the predictions of a model that assumes aversion to inequitable outcomes (Fehr & Schmidt 1999). This simple model makes it possible to develop alternative predictions in our complex game. We do not argue that inequity aversion is the most realistic model of other-regarding preferences. Instead, the goal of this Section is to show that systematic differences from the selfish model are expected for a wide range of parameter values of one of the most influential social preferences models.

With inequity aversion, if teachers are prosocial, a higher fixed wage can actually reduce bribery. To see this, recall that students of sufficient ability do best when they exert effort with a teacher who does not solicit bribes. Teachers receiving a fair wage can reduce inequality with respect to these students by not soliciting bribes. This mechanism can be a powerful motive for teachers. Whether it actually is depends on the details, namely teacher wages, the extent to which teachers are inequity-averse, and the distribution of student ability. This complex mix is why the general model with prosocial preferences is intractable. We, therefore, restrict the predictions for the case of other-regarding preferences to the constellation of parameters used in the experiment.

Specifically, we derive predictions by assuming that teachers exhibit inequity aversion based on the utility function (Fehr & Schmidt 1999),

$$U(x_i, x_{-i}) = x_i - \alpha \sum_{j \neq i} \max(x_j - x_i, 0) - \beta \sum_{j \neq i} \max(x_i - x_j, 0),$$

where x_i is the payoff of a focal teacher, x_{-i} is a vector of payoffs for the other teacher and the students, α measures aversion to disadvantageous inequality, and β aversion to advantageous inequality.[6] Appendix C reports the derivation of the predictions. Figure 2 shows how the set of equilibria in the different parts of the two treatments depend on α and β. Equilibria are shown in different colors as a function of aversion to disadvantageous inequality (α) and advantageous inequality (β). The dots in each figure reports different estimated values of α and β reported in the literature (Goeree & Holt 2000, Blanco et al. 2011, Beranek et al. 2015).

——————— Figure 2 about here ———————

Predictions clearly depend on both the incentive scheme and the prosocial motives among teachers. Under the *low fixed wage* we implemented for the first 15 periods of both treatments (panel (a)), the unique equilibrium involves both teachers soliciting bribes. Intuitively, when teachers are paid little, in particular less than students, bribes actually reduce inequality, and in this sense inequity aversion among teachers actually supports soliciting bribes. Under the *high fixed wage* we implemented in the second 15 periods of our fixed-wage treatment (panel (b)), most empirical estimates of α and β instead point toward a unique equilibrium in which neither teacher solicits bribes. In this case, teachers typically receive more than students, and thus soliciting bribes increases inequality in two ways. Namely, bribes consist of direct transfers from students to teachers, and bribes also impoverish students further by reducing the value of a diploma. The resulting inequalities create disutilities for teachers, and this is how an unconditional pay raise for prosocial teachers can actually break a corrupt system.

Interestingly, under the *piece rate* we implemented in the second 15 periods of our piece-rate treatment (panel (c)), most empirical estimates of α and β lead to two possible

[6]In this analysis we continue to make the simplifying assumption that students are selfish. Students' choices have a lesser impact on the distribution of payoffs than teachers' choices. For a range of parameters that includes a series of estimates in the literature, the predictions listed above remain the same when inequity aversion among students is introduced. See Appendix C for the details.

equilibria, one with neither teacher soliciting bribes and the other with both teachers soliciting bribes. The first is an equilibrium because by deviating a teacher attracts only a minority of the students and is worse off compared to the other teacher and to most of the students. The second is an equilibrium because by unilaterally deviating, the teacher attracts most of the students, which creates a high level of advantageous inequality towards all the students and the other teacher.

This analysis shows that a fair wage with reasonable levels of inequity aversion can break corruption in our experiment. Based on this, Hypothesis H3 will be rejected in favor of the alternative hypothesis that in treatment *FW*, the fraction of bribes solicited (and paid) significantly decreases in the second part compared to the first part.[7]

Results

Teachers' behavior and overall frequency of bribing

The violin plots of Figure 3 illustrate the distribution of the fraction of bribes solicited by the teachers. Part 1 describes the first 15 periods of play (panel (a)). Here, bribery levels exceed 90% for the majority of the groups and do not differ across the two treatments (Wilcoxon rank sum test $p = 0.426$).[8] However, in part 2 (panel (b)), the fraction of teachers soliciting bribes is significantly lower in the piece-rate treatment compared to the fixed-wage treatment (Wilcoxon rank sum test $p < 0.001$). While bribery levels remain high in the fixed-wage treatment (Wilcoxon signed rank test $p = 0.750$), they significantly drop between part 1 and part 2 in the piece-rate treatment (Wilcoxon signed rank test $p < 0.001$). This effect over time is also depicted in panel (c) illustrating the fraction of teachers soliciting bribes over periods while collapsing across groups.

[7] The foregoing analysis is based on an equilibrium analysis of the stage-game. Naturally, repeated play of a stage-game equilibrium will be an equilibrium of the repeated game. From the behavioral literature, it is well known that people often aim to exploit the repeated character of a game. An interesting behavioral prediction about the solicitation of bribes can be derived if teachers are "imperfect conditional cooperators" who match others' contributions only partly. Fischbacher & Gächter (2010) show that this is an important motivation that explains why in many finitely repeated public good games subjects start by aiming for cooperation which decays over time. Applied to our setup, this approach would suggest that in part 2 of treatment *PR* teachers may continue to cooperate by soliciting bribes, but that such attempts unravel over time.

[8] All pairwise comparisons in the paper are conducted using groups' averages as the unit of observation.

Taken together, in the first part around 90% of teachers solicit bribes, independent of the treatment. In the second part, the fraction of teachers soliciting bribes remains high after a six-fold pay-raise in the fixed-wage treatment, but drops to 50% after the introduction of the piece-rate regime. Hence, the results confirm the model's predictions by not rejecting H1 & H3, while rejecting H2 & H4.

———————— Figure 3 about here ————————

To corroborate these findings, we estimated a series of linear probability models on individual level data. The results are displayed in columns 1 to 3 of Table 1. Models 1, 2, and 3 use a dummy variable taking value 1 if the teacher solicits bribes as a dependent variable. All three models consider heteroskedasticity-robust standard errors clustered at the group level. These models permit us to better explore how the decision to solicit bribes changes over time while controlling for demographics (age and gender), income, and whether subjects have been ever asked to pay a bribe in their life.[9]

Results for the first part indicate that: (i) the probability of teachers soliciting bribes does not differ between treatments, neither in the levels (d(*piece-rate*) in Models 2 and 3), nor in the evolution over time (*Period*×d(*piece-rate*) in Models 2 and 3); (ii) the probability of bribe extraction increases over periods (*Period* in Models 2 and 3). In the second part, however, (i) the probability of teachers soliciting bribes is significantly lower in *PR* compared to *FW* (d(*piece-rate*)×d(*part 2*)) and (ii) this difference between treatments increases over periods (*Period*×d(*piece-rate*) in Model 2 and *Period*×d(*piece-rate*)×d(*part 2*) in Model 3). These results clearly favor the predictions derived with selfish preferences over the ones derived with inequity aversion. In particular, we do not observe any reduction in the fraction of bribes solicited in the second part of the *FW* treatment.

[9] Logit and Probit regressions with errors clustered at the group level provide similar results. The differences compared to the linear probability model are in the interaction term *Period*×d(*piece-rate*)×d(*part 2*) of Model 3 that becomes non-significant and in the control variable d(*asked to pay a bribe*) that in some of the regression becomes significant at 10% level.

——————— Table 1 about here ———————

Besides analyzing the fraction of bribes solicited by the teachers, we look at the overall occurrence of bribery. Note that the overall frequency of bribery depends on both the teachers' and the students' decisions, as bribery only occurs if the teacher *solicits* bribes and the student *pays* the bribe. Figure D.1 in appendix D shows the overall frequency of bribes paid in the group by part and by treatment and the fraction of bribes paid over periods.

Columns 4-6 of Table 1 report a series of linear probability models that are estimated using individual level data. The dependent variable is a dummy taking value 1 if a student paid a bribe in that period. Explanatory variables are the same as in Table 1. Results confirm the observed effects of teachers' demand for bribes: (i) in the first part, overall levels of bribery do not differ across treatments and become more frequent over periods; (ii) in the second part, however, significant treatment differences emerge, with lower levels of bribery in the *PR* treatment—this difference even increases over periods. Among the control variables, participants' experience with bribery significantly positively affected bribe payments, translating into a 7 percentage points increase.[10]

Students' individual behavior

Next, we test whether the students' behavior aligns with the theoretical predictions. Here, we compare the actual rate of bribes observed in the group with the predicted rate calculated assuming that students follow the equilibrium strategy (based on selfish preferences) given their effort costs and the teachers' actual choices. As can be seen in Figure 4, aggregated student behavior largely matches the equilibrium predictions. While in the first part the actual fraction of bribes paid is slightly lower than the predicted

[10] An important empirical question is how robust is the effect of piece-rate incentives to changes in the distribution of effort costs. From a theoretical perspective, our proposition highlights that predictions are indeed sensitive to the distribution of effort costs. From an empirical perspective, the effort cost of intermediate students can be an important parameter. If this, for instance, would be higher and thereby closer to the effort cost of the bad students, intermediate students could be closer to indifference, impacting the speed at which the piece-rate intervention would reduce bribe solicitation.

fraction in most of the groups (panel (a)). In the second part, the observed fractions are closer to the predicted fractions with the observations lying closely around the diagonal (panel (b)). Overall, the likelihood to observe students playing the strategy predicted by the model is generally high in both parts of the experiment (78% in part 1 and 84% in part 2).

————— Figure 4 about here —————

Alternative motives for the students' choices may exist. A prime candidate is moral concerns that curb students' willingness to pay bribes (Köbis et al. 2016). Additional analyses reported in the appendix D estimated how frequently students deviated from equilibrium predictions due to moral concerns. The comparison of the equilibrium profile to the best strategy profile in which students do not pay bribes when solicited provides very little support for the presence of such moral concerns in the current set-up.

Discussion

Our paper studies the dynamics of bribery in education by using a relatively simple game that captures the key strategic features of the interaction. We model and experimentally confirm that letting students choose between schools can give them leverage to change systems of corruption when teachers' incentives are properly designed. We used a lab experiment to establish causal effects given that dynamics in the field, for example, changes in the reputational value of a degree, would take years to materialize. We increased external validity by creating a robust norm of widespread corruption of teachers soliciting bribes before testing anti-bribery interventions (Muthukrishna et al. 2017). Surprisingly, despite teachers retaining discretionary power to solicit bribes, their inclination to do so diminishes significantly upon the implementation of the new piece-rate system. Moreover, we examined bribery schemes between students and teaching staff, for which other studies have shown high correlations between lab and field measures (Armantier & Boly

2013). Our results thus serve as a proof of concept for the idea that a piece-rate scheme can reduce bribery in education by shifting power to students. Prior to implementing this approach in practice, it is essential to carry out a customized evaluation of costs and benefits, considering diverse factors concerning the specific community involved, such as the significance of having a teacher stationed in a remote area.

Naturally, there are many ways in which the situation in practice can differ from the one that we studied in the experiment. Here, we note that Armantier & Boly (2013) provides supportive evidence that lab and field data on bribery practices in the education sector largely overlap. They compare how graders of exams respond to bribing requests in the lab in Canada as well as both in the lab and in the field in Burkina Faso. Not only the direction but also the magnitude of their treatment effects are statistically indistinguishable across the three environments. These findings bolster our conjecture that our evidence may extrapolate to the field and that compared with a fixed-wage increase, enhancing wages through a piece-rate regime may be a more promising avenue to reduce bribery.

Limitations and Extensions of the Model

Extensions of our simple model can provide insights into limitations and boundary conditions for the intervention to succeed. First, under the assumption that players are selfish, which is supported by the experimental data, the predictions for the fixed-wage regime remain the same in the extensions that we consider here. That is, teachers will continue to cash in on higher wages without changing their rent-seeking behaviors. The predictions of the piece-rate regime will vary. Sometimes the effect of competition is dampened, but theoretically, the piece-rate regime always performs at least as well as the fixed-wage regime (see Appendix A for formal analyses of these points).

Second, we consider how the predictions change when more or less than two schools compete with each other (see for more details appendix A.3). With one school, a piece-rate regime does not outperform a fixed-wage regime, unless it offers substantial incentives to open a new competing school. With more than two schools, teachers will choose to diversify, with some teachers choosing to abstain from bribes and targeting good students and other teachers choosing to solicit bribes and targeting bad students. Although corruption rates drop less steeply, it may actually lead to a Pareto-improvement compared to the case where all teachers do not solicit bribes. In the latter case, bad students suffer because they were used to getting a discredited diploma without working hard and are now forced to opt out, which provides them with a lower payoff. As a consequence, the piece-rate mechanism has the potential to create reputable schools and universities in highly corrupt contexts—which can serve as possible role models—motivating other schools to follow suit (this is in line with recent anti-corruption policy approaches focusing on so-called positive deviance and "islands of integrity" see for example, Jackson & Köbis 2018, Zuniga 2018).

Third, given that teachers cannot handle an unlimited number of students, it is important to consider the implications of an upper-bound on the maximum class-size (see appendix A.5). As long as the ratio between maximum class size and student population is small, predictions do not change. However, the incentives to compete for good students decrease when the restrictions on class-sizes are more stringent. In such cases, teachers may want to divide the market so that one teacher targets good students and does not solicit bribes, while the other teacher targets bad students and solicits bribes. As a companion intervention to limit the total number of students applying to a single school, one can allow only competition within districts composed of a small number of schools that can accommodate the students of the district.

Fourth, we consider what happens when allowing teachers to choose the size of the bribes they solicit (see appendix A.2). In our baseline model, the size of the bribes is fixed. Teachers merely decide whether or not to solicit bribes. Although continuous bribes complicate the strategic decision, they do not affect the main predictions. Akin to the base-line model, sufficient subsidies lead both teachers to abstain from soliciting bribes.[11]

[11] Another variation would be to replace the free choice to pay the bribe with mandatory payment. In that case, a teacher who is soliciting bribes forces students to pay the bribe. In the theoretical model, mandatory bribery increases the likelihood of better students paying bribes and, at the same time, gives students a stronger incentive to avoid bribing classrooms. Other than that, the theoretical predictions

Overcoming implementation challenges for policy makers

Policymakers hoping to reduce bribery in education by reforming teacher salaries should consider mechanisms that concurrently shift power to those directly on the receiving end of the damage corruption causes. Both our model and experiment show that one way to do this is to allow students to choose and to pay teachers according to student choices. This mechanism begins with the simple fact that the externalities of obtaining a degree are substantial compared to the effort required to obtain it so most of the students would prefer to study hard for a more valuable degree in lieu of paying bribes for a less valuable degree. The mechanism then ensures that these students have the power to incentivise educators to supply bribery-free schools.

Like any new approach to tackling entrenched corruption, using a piece rate to incentivise teacher integrity could bring important challenges in the field. The results reported here are of a lab experiment. Any policy recommendation based on such findings should be done with care. Results suggest the piece-rate mechanism may have the potential to create reputable schools and universities in highly corrupt contexts—which can serve as possible role models. That said, before implementing such a mechanism, a careful welfare analysis, considering the local set-up and related stakeholders, must be conducted to assess if the potential societal profits outweigh the costs of implementation.

Several infrastructural preconditions need to be in place for the proposed market for integrity to emerge. First, students/parents need to be able to choose, at relatively low costs, between different schools. Hence, in rural regions with sparse school density the proposed piece-rate mechanism may fail because each school can essentially act as a monopolist.

Second, the proposed model rests on the assumption that students—or their parents—have the desire and ability to choose the "right" school. Whether or not this is the case is ultimately an empirical matter. Note that in Chile, Chumacero et al. (2011) find that parents pay attention to quality when choosing schools. Still, information asymmetries, e.g., between families with different socio-economic status or systematic biases in the school

of the baseline model remain the same.

choice based on non-corruption-related issues, could potentially undermine an informed school choice and hence reduce the effectiveness of the piece-rate scheme (Nambissan & Ball 2010). Facilitating the distribution and availability of relevant information within the target population helps to harness the potential of piece-rate schemes. Engagement of parents in school committees can help to disseminate relevant information (Duflo et al. 2015, Wood & Antonowicz 2011).

Third, policymakers need to properly calibrate the size of the incentives. Our model predicts that a piece-rate regime will succeed only if the subsidy is sufficiently high to trigger actual student competition. The case of Chile shows that a modest subsidy per student may fail to be effective.

Finally, new types of fraud, especially if monitoring is imperfect, can arise. For example, teachers and schools might attempt to circumvent the payment scheme. They might come up with fraudulent ways to get the piece-rate, such as the admission of unqualified students or fake enrollment of students (see for some challenges Borcan et al. 2014). A calibration of the piece-rate to the respective educational system and the adoption of modern payment technology to administer the piece-rate can help to reduce such pitfalls (Hanna 2017). The challenge of ineffective monitoring can be overcome by involving parents and students. For example, Duflo et al. (2012) let students monitor the teachers' attendance by taking a picture of the teacher and the other students at the start and end of a given school day. With incentives for teachers being contingent on these pictures, this intervention in turn successfully contributed to the reduction of absenteeism. Combining the piece-rate scheme with modern payment technology and bottom-up monitoring is thus advisable to harness the full potential of a piece-rate-based intervention.

Conclusions

Although bribery in education brings immense societal costs (Heyneman et al. 2008), and even though recent policy work has emphasized the need for behavioural insights (OECD 2018), surprisingly little behavioural research on the topic exists. One popular approach to counteract the negative effects of bribery is fixed public salary increases. Our theoretical model and the results of a large pre-registered lab experiment suggest that such unconditional salary increases are ineffective in reducing bribery—plausibly because the effectiveness of such wage increases requires effective punishment institutions to provide realistic deterrence. However, in high corruption contexts where anti-corruption reforms are most urgent, reliable punishment institutions are lacking as impunity emerges.

As a new alternative for such systems riddled by bribery, we propose a market design that helps to empower students and their parents to reward integrity of teaching staff by voting with their feet. Instead of focusing on the stick—efforts to introduce effective punishment regimes in highly corrupt contexts have largely failed—this approach focuses on the carrot—reshaping incentives for teachers to offer bribe-free classrooms. Our proposed piece-rate scheme does not rely on "honest principals". Instead, self-interested teachers will understand that it is in their best interest to opt for integrity.

The results of the theoretical model and the lab experiment provide first empirical support for the effectiveness of the piece-rate regime—although not eliminating bribery altogether, it substantially reduced its occurrence. The effectiveness of the piece-rate intervention bears additional weight when considering that it was introduced after a widespread norm of corruption was established. Another noteworthy feature of the approach is that the fixed-wage intervention and the piece-rate intervention are budget balanced. Our results thus suggest that reshaping the incentives of a salary policy can potentially contribute to the emergence of market forces that run counter to the incentives of corruption. When implemented wisely, competition can create a market for integrity.

References

Abbink, K. (2002), Fair salaries and the moral costs of corruption, Technical report, Bonn Econ Discussion Papers.

Abbink, K. & Hennig-Schmidt, H. (2006), 'Neutral versus loaded instructions in a bribery experiment', *Experimental Economics* **9**(2), 103–121.

Akerlof, G. A. & Yellen, J. L. (1990), 'The fair wage-effort hypothesis and unemployment', *The Quarterly Journal of Economics* **105**(2), 255–283.

An, W. & Kweon, Y. (2017), 'Do higher government wages induce less corruption? cross-country panel evidence', *Journal of Policy Modeling* **39**(5), 809–826.

Armantier, O. & Boly, A. (2011), 'A controlled field experiment on corruption', *European Economic Review* **55**(8), 1072–1082.

Armantier, O. & Boly, A. (2013), 'Comparing corruption in the laboratory and in the field in burkina faso and in canada', *The Economic Journal* **123**(573), 1168–1187.

Barr, A., Lindelow, M. & Serneels, P. (2009), 'Corruption in public service delivery: An experimental analysis', *Journal of Economic Behavior & Organization* **72**(1), 225–239.

Barr, A. & Serra, D. (2009), 'The effects of externalities and framing on bribery in a petty corruption experiment', *Experimental Economics* **12**(4).

Becker, G. S. & Stigler, G. J. (1974), 'Law Enforcement, Malfeasance, and Compensation of Enforcers', *The Journal of Legal Studies* **3**(1), 1–18.

Bennett, V. M., Pierce, L., Snyder, J. A. & Toffel, M. W. (2013), 'Customer-driven misconduct: How competition corrupts business practices', *Management Science* **59**(8), 1725–1742.

Beranek, B., Cubitt, R. & Gächter, S. (2015), 'Stated and revealed inequality aversion in three subject pools', *Journal of the Economic Science Association* **1**(1), 43–58.

Blanco, M., Engelmann, D. & Normann, H.-T. (2011), 'A within-subject analysis of other-regarding preferences', *Games and Economic Behavior* **72**(2), 321–338.

Borcan, O., Lindahl, M. & Mitrut, A. (2014), 'The impact of an unexpected wage cut on corruption: Evidence from a "xeroxed" exam', *Journal of Public Economics* **120**, 32–47.

Borcan, O., Lindahl, M. & Mitrut, A. (2017), 'Fighting corruption in education: What works and who benefits?', *American Economic Journal: Economic Policy* **9**(1), 180–209.

Chaudhury, N., Hammer, J., Kremer, M., Muralidharan, K. & Rogers, F. H. (2006), 'Missing in action: teacher and health worker absence in developing countries', *Journal of Economic Perspectives* **20**(1), 91–116.

Chumacero, R. A., Gómez, D. & Paredes, R. D. (2011), 'I would walk 500 miles (if it paid): Vouchers and school choice in chile', *Economics of Education Review* **30**(5), 1103–1114.

Conti, G., Heckman, J. J. & Urzua, S. (2010), 'The education-health gradient', *American Economic Review* **100**(2), 234–38.

Cooray, A. & Schneider, F. (2016), 'Does corruption promote emigration? an empirical examination', *Journal of Population Economics* **29**(1), 293–310.

de Ree, J., Muralidharan, K., Pradhan, M. & Rogers, H. (2017), 'Double for Nothing? Experimental Evidence on an Unconditional Teacher Salary Increase in Indonesia*', *The Quarterly Journal of Economics* **133**(2), 993–1039.
URL: *https://doi.org/10.1093/qje/qjx040*

Di Tella, R. & Schargrodsky, E. (2003), 'The role of wages and auditing during a crackdown on corruption in the city of buenos aires', *Journal of Law and Economics* **46**(1), 269–292.
URL: *https://EconPapers.repec.org/RePEc:ucp:jlawec:y:2003:v:46:i:1:p:269-92*

Dolan, J., Winthrop, R., Golden, A. & Ndaruhutse, S. (2012), *Building effective teacher salary systems in fragile and conflict-affected states*, CfBT Education Trust.

Dorrough, A. R., Köbis, N., Irlenbusch, B., Shalvi, S. & Glöckner, A. (2023), 'Conditional bribery: Insights from incentivized experiments across 18 nations', *Proceedings of the National Academy of Sciences* **120**(18), e2209731120.

Duflo, E., Dupas, P. & Kremer, M. (2015), 'School governance, teacher incentives, and pupil–teacher ratios: Experimental evidence from kenyan primary schools', *Journal of Public Economics* **123**, 92–110.

Duflo, E., Hanna, R. & Ryan, S. P. (2012), 'Incentives work: Getting teachers to come to school', *American Economic Review* **102**(4), 1241–78.

Falk, A. & Heckman, J. J. (2009), 'Lab experiments are a major source of knowledge in the social sciences', *Science* **326**(5952), 535–538.
URL: *https://science.sciencemag.org/content/326/5952/535*

Fehr, E. & Schmidt, K. M. (1999), 'A theory of fairness, competition, and cooperation', *The Quarterly Journal of Economics* **114**(3), 817–868.

Fischbacher, U. (2007), 'z-tree: Zurich toolbox for ready-made economic experiments', *Experimental Economics* **10**(2), 171–178.

Fischbacher, U. & Gächter, S. (2010), 'Social preferences, beliefs, and the dynamics of free riding in public goods experiments', *American Economic Review* **100**(1), 541–56.

Fisman, R. & Golden, M. A. (2017), *Corruption: What everyone needs to know*, Oxford University Press.

Foltz, J. D. & Opoku-Agyemang, K. A. (2015), 'Do higher salaries lower petty corruption? a policy experiment on west africa's highways', *Unpublished Working Paper, University of Wisconsin-Madison and University of California, Berkeley* .

Friedman, M. (1955), The role of government in education, *in* R. A. Solo, ed., 'Economics and the Public Interest', Rutgers University Press, New Brunswick, NJ, pp. 123–144.

Friedman, M. (1962), *Capitalism and freedom*, University of Chicago press.

Global Partnership for Education (2020), Global Projects Portfolio, Technical report.

Goeree, J. K. & Holt, C. A. (2000), 'Asymmetric inequality aversion and noisy behavior in alternating-offer bargaining games', *European Economic Review* **44**(4), 1079–1089.

Hallak, J. & Poisson, M. (2007), *Corrupt schools, corrupt universities: What can be done?*, International Institute for Education Planning Paris.

Hanna, R. (2017), 'Technology beats corruption', *Science* **355**(6322), 244–245.

Heyneman, S. P., Anderson, K. H. & Nuraliyeva, N. (2008), 'The Cost of Corruption in Higher Education', *Comparative Education Review* **52**(1), 1–25.

Jackson, D. & Köbis, N. C. (2018), Anti-corruption through a social norms lens, Technical report, Chr. Michelsen Institute (U4 Anti-Corruption Resource Center - Issue 7).

Jacquemet, N. (2005), Corruption as betrayal: Experimental evidence on corruption under delegation, Technical report, Groupe d'Analyse et de Théorie Economique Lyon St-Étienne (GATE Lyon).

Köbis, N. C., van Prooijen, J.-W., Righetti, F. & Van Lange, P. A. (2016), 'Prospection in individual and interpersonal corruption dilemmas.', *Review of General Psychology* **20**(1), 71.

Köbis, N., Starke, C. & Rahwan, I. (2022), 'The promise and perils of using artificial intelligence to fight corruption', *Nature Machine Intelligence* **4**(5), 418–424.

Lambsdorff, J. G. & Fink, H. (2006), Combating corruption in colombia: Perceptions and achievements, Passauer Diskussionspapiere, Volkswirtschaftliche Reihe V-44-06, University of Passau, Faculty of Business and Economics.
URL: *https://EconPapers.repec.org/RePEc:zbw:upadvr:v4406*

Luca, M., Rooney, P. & Smith, J. (2017), 'The impact of campus scandals on college applications', *Harvard Business School NOM Unit Working Paper* **16-137**.

Mavisakalyan, A. & Meinecke, J. (2016), 'The labor market return to academic fraud', *European Economic Review* **82**, 212–230.
URL: *https://www.sciencedirect.com/science/article/pii/S0014292115001828*

Mungiu-Pippidi, A. (2017), 'The time has come for evidence-based anticorruption', *Nature Human Behaviour* **1**.

Munro, C. & Kirya, M. (2020), Values education for public integrity - What works and what doesn't, Technical report, Chr. Michelsen Institute (U4 Anti-Corruption Resource Center - Issue 8).

Muthukrishna, M., Francois, P., Pourahmadi, S. & Henrich, J. (2017), 'Corrupting cooperation and how anti-corruption strategies may backfire', *Nature Human Behaviour* **1**(7), 0138.

Muthukrishna, M. & Henrich, J. (2019), 'A problem in theory', *Nature Human Behaviour* **3**(3), 221–229.

Nambissan, G. B. & Ball, S. J. (2010), 'Advocacy networks, choice and private schooling of the poor in india', *Global Networks* **10**(3), 324–343.

OECD (2016), *Education in Colombia*, OECD Publishing.
URL: *https://www.oecd-ilibrary.org/content/publication/9789264250604-en*

OECD (2018), *Behavioural Insights for Public Integrity: Harnessing the Human Factor to Counter Corruption*, OECD Publishing.

Rauch, J. E. & Evans, P. B. (2000), 'Bureaucratic structure and bureaucratic performance in less developed countries', *Journal of Public Economics* **75**(1), 49–71.

Rose-Ackerman, S. & Palifka, B. J. (2016), *Corruption and government: Causes, consequences, and reform*, Cambridge University Press.

Rothstein, B. (2011), 'Anti-corruption: the indirect 'big bang'approach', *Review of International Political Economy* **18**(2), 228–250.

Shapiro, C. & Stiglitz, J. E. (1984), 'Equilibrium unemployment as a worker discipline device', *The American Economic Review* **74**(3), 433–444.

Shleifer, A. (2004), 'Does competition destroy ethical behavior?', *American Economic Review* **94**(2), 414–418.

Stephenson, M. C. (2020), 'Corruption as a self-reinforcing trap: Implications for reform strategy', *The World Bank Research Observer* **35**(2), 192–226.

Sutter, M., Haigner, S. & Kocher, M. G. (2010), 'Choosing the carrot or the stick? endogenous institutional choice in social dilemma situations.', *Review of Economic Studies* **77**(4), 1540 – 1566.
URL: *https://search.ebscohost.com/login.aspx?direct=true&db=bah&AN=52904214&site=ehost-live&scope=site*

Svensson, J. (2005), 'Eight questions about corruption', *Journal of Economic Perspectives* **19**(3), 19–42.

Tanzi, V. (1998), 'Corruption Around the World: Causes, Consequences, Scope and Cures', *IMF Staff Papers* **45**(4), 559–594.

Transparency International (2013), *Global corruption barometer*, Transparency International Berlin.

Treisman, D. (2000), 'The causes of corruption: a cross-national study', *Journal of public economics* **76**(3), 399–457.

Van Rijckeghem, C. & Weder, B. (2001), 'Bureaucratic corruption and the rate of temptation: do wages in the civil service affect corruption, and by how much?', *Journal of Development Economics* **65**(2), 307–331.

Van Veldhuizen, R. (2013), 'The influence of wages on public officials' corruptibility: A laboratory investigation', *Journal of Economic Psychology* **39**, 341–356.
URL: *http://dx.doi.org/10.1016/j.joep.2013.09.009*

Voors, M., Turley, T., Bulte, E., Kontoleon, A. & List, J. A. (2018), ‘Chief for a day: Elite capture and management performance in a field experiment in sierra leone’, *Management Science* **64**(12), 5855–5876.
URL: *https://doi.org/10.1287/mnsc.2017.2866*

Wood, J. & Antonowicz, L. (2011), *Fighting Corruption in the Education Sector: Methods Tools, and Good Practices. United Nations Development Programme. 2011*, United Nations Development Programme.

Zuniga, N. (2018), The impact of ’islands of integrity’, Technical report, Chr. Michelsen Institute (U4 Anti-Corruption Resource Center - Helpdesk Answer 1).

Figure 1: Equilibria in the *piece-rate* regime as a function of $\mathcal{F}(a+b)$ and s/b (the example assumes $\mathcal{F}(1-\pi_{out}) = 0.95$ and $\frac{E_b}{N} = 0.5$).

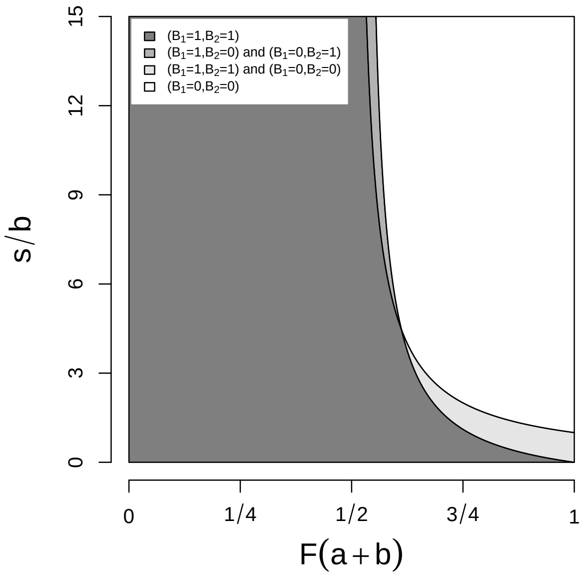


Figure 2: Equilibria in part 1 and part 2 of the two treatments when teachers are inequity averse with Fehr & Schmidt (1999) preferences. The points represent estimated parameters found in the literature: GH refers to the parameters for the proposer and responder estimated in Goeree & Holt (2000), BEN refers to the parameters estimated in Blanco et al. (2011), and BCG refers to the parameters estimated by Beranek et al. (2015) in three different samples.

(a) FW and PR - Part 1 (b) FW - Part 2 (c) PR - Part 2

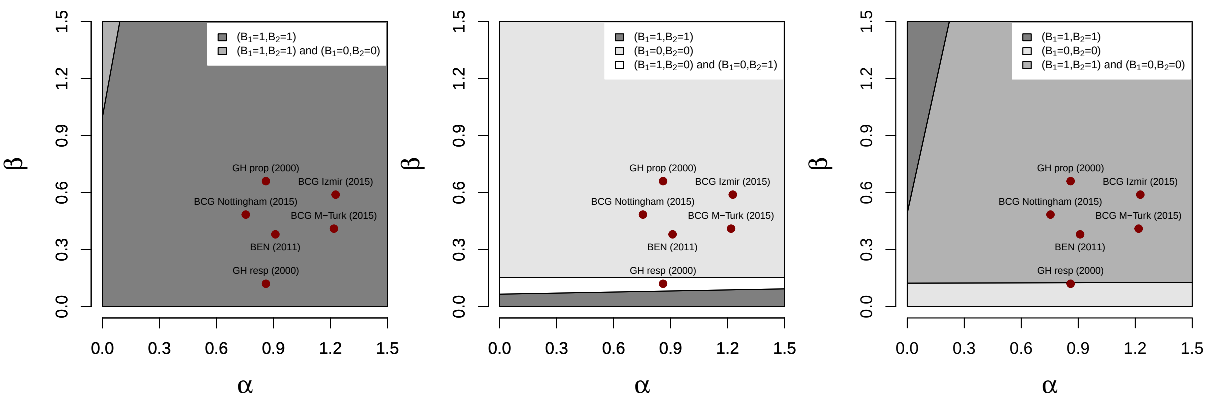
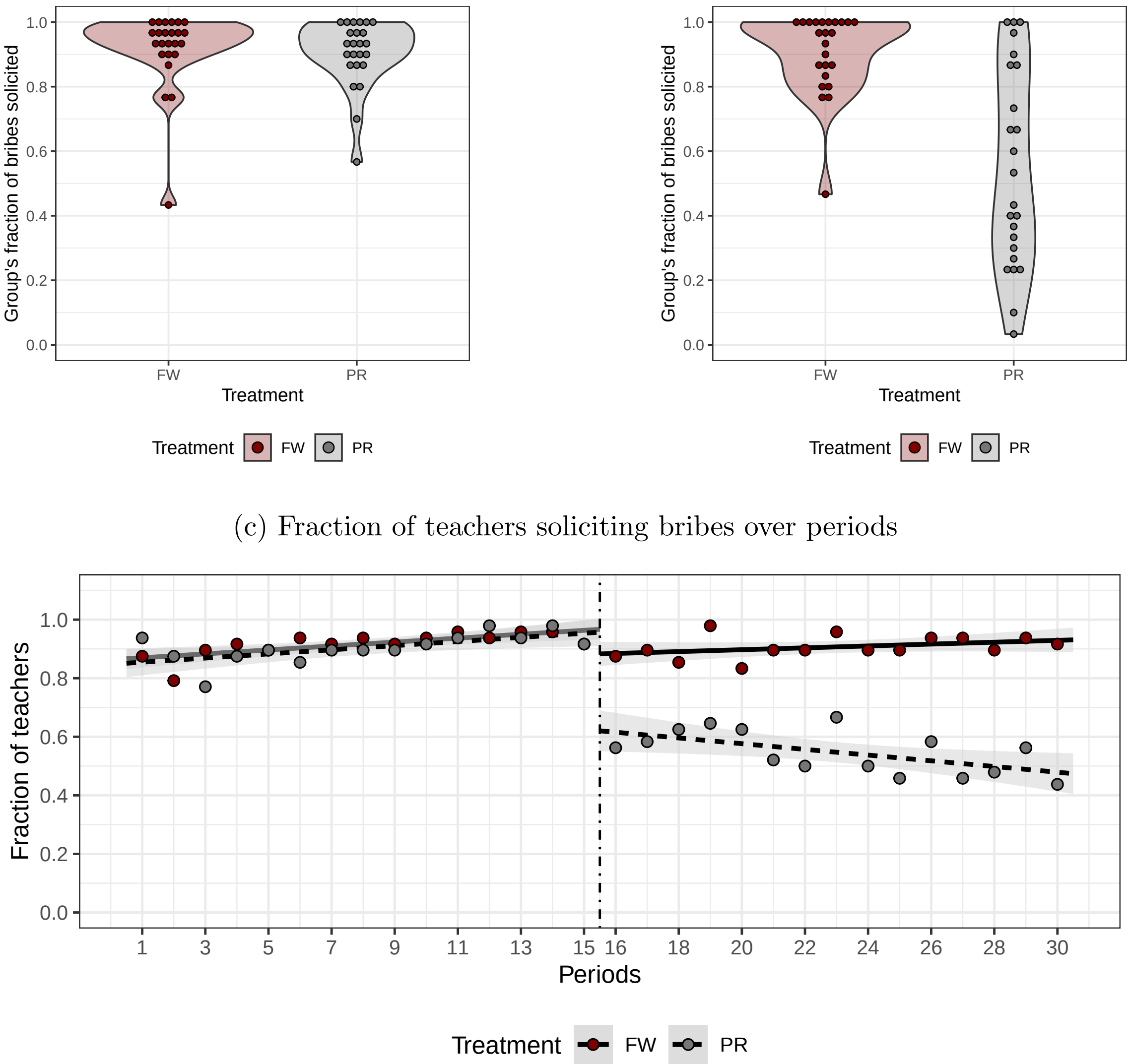


Figure 3: Groups' fractions of bribes solicited in part 1 (a) and in part 2 (b) and fraction of teachers soliciting bribes over periods (c)



(a) Part 1

(b) Part 2

(c) Fraction of teachers soliciting bribes over periods

Figure 4: Actual vs. predicted rate of bribes paid in the group. Predictions are made assuming that students follows the equilibrium strategy given their effort costs and the actual choices of the teachers.

(a) Part 1

(b) Part 2

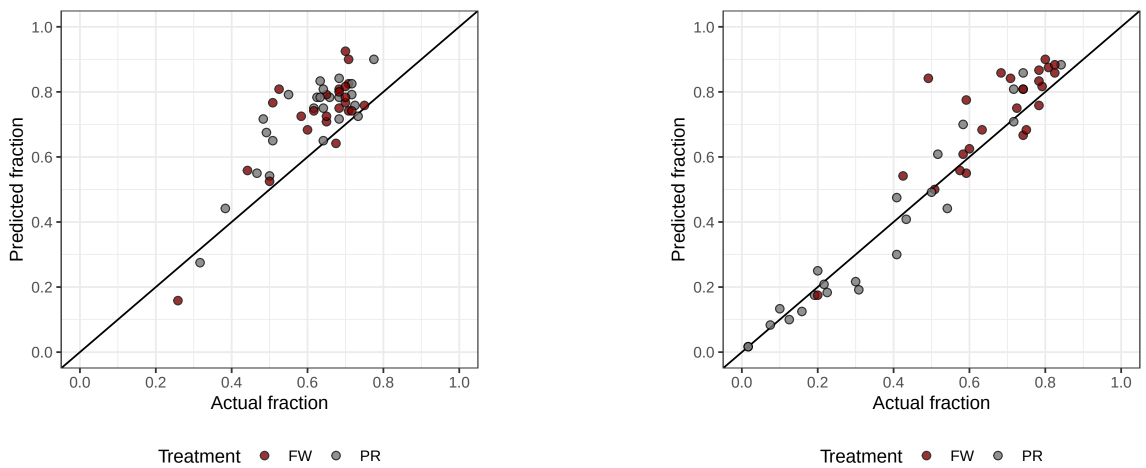


Table 1: Linear probability models (Probability that bribes are solicited and paid). Heteroscedasticity robust s.e. are reported in parentheses. Errors are clustered at group level.

	Bribes solicited			Bribing rates		
	Mod. 1	Mod. 2	Mod. 3	Mod. 4	Mod. 5	Mod. 6
	Part 1 data	Part 2 data	All data	Part 1 data	Part 2 data	All data
(Intercept)	0.882***	0.769***	0.820***	0.554***	0.775***	0.632***
	(0.174)	(0.226)	(0.173)	(0.098)	(0.108)	(0.090)
Period	0.006**	0.003	0.006**	0.009***	0.005°	0.009***
	(0.002)	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)
d(*piece-rate*)	-0.013	-0.261***	-0.014	-0.016	-0.211***	-0.017
	(0.041)	(0.073)	(0.043)	(0.040)	(0.060)	(0.040)
Period×d(*piece-rate*)	0.001	-0.013*	0.001	-0.001	-0.011°	-0.001
	(0.003)	(0.006)	(0.003)	(0.004)	(0.006)	(0.004)
d(*part 2*)	—	—	0.011	—	—	0.066*
	—	—	(0.020)	—	—	(0.033)
d(*part 2*)×*Period*	—	—	-0.003	—	—	-0.004
	—	—	(0.002)	—	—	(0.003)
d(*part 2*)×d(*piece-rate*)	—	—	-0.246***	—	—	-0.193**
	—	—	(0.059)	—	—	(0.059)
d(*part 2*)×*Period*×d(*piece-rate*)	—	—	-0.013*	—	—	-0.010
	—	—	(0.007)	—	—	(0.007)
Age	0.005	0.014	0.009	-0.002	-0.010°	-0.006
	(0.008)	(0.012)	(0.008)	(0.005)	(0.006)	(0.005)
d(*male*)	-0.053	-0.049	-0.051	-0.019	0.006	-0.007
	(0.042)	(0.059)	(0.047)	(0.021)	(0.023)	(0.018)
Income	-0.031°	-0.045	-0.038°	0.011	0.013	0.012
	(0.016)	(0.032)	(0.022)	(0.011)	(0.016)	(0.011)
d(*asked to pay a bribe in the past*)	-0.102	-0.111	-0.106	0.061*	0.074*	0.068**
	(0.076)	(0.085)	(0.076)	(0.027)	(0.034)	(0.024)
n	1410	1410	2820	5700	5700	11400

Signif. codes: $0 < {***} \leq 0.001 < {**} \leq 0.01 < {*} \leq 0.05 < {\circ} \leq 0.10$

A. Theory: base model and extensions

A.1 Proof of the main proposition

Proof. We solve the game by backward induction. We start with the shortest sub-game.

Stage 3: Choice whether or not to pay the bribe. We first consider the shortest sub-game in which a student chooses a teacher who solicits bribes. In this case, it is profitable to pay the bribe when the effort saved is greater than the bribe plus the damage of reputation produced, i.e., $c_j > b + \frac{1}{n}a$, or when

$$n > \frac{a}{c_j - b}.$$

So, good students—with $c_j \leq b + \frac{1}{N}a$—do not pay the bribe for any n; bad students—with $c_j > b + a$—pay the bribe for all n; and average students pay the bribe if the school is big enough. The number $n^*(c_j) = \left\lfloor \frac{a}{c_j - b} \right\rfloor$ is the maximum size of the school for which it is not profitable to pay for a student with effort cost c_j.

Stage 2: Choice of the school. We now consider the sub-games in which students choose which school to attend if any. Three sub-games exist:

- **Both teachers do not solicit bribes.** In this case, students have two options: (i) to go to school and study hard to get the diploma, thereby receiving a payoff of $1 - c_j$; or (ii) to not go to school and get a payoff of π_{out}. Therefore, students with $c_j > 1 - \pi_{out}$ prefer not to go to school and students with $c_j \leq 1 - \pi_{out}$ prefer to go to school. Since students who prefer to go to school are indifferent between the two teachers, we assume that they are randomly split between them.

- **One teacher solicits bribes and the other teacher does not.** All students use the same cutoff strategy. This means that all students with a c_j at least as large as the cutoff c^* choose the teacher soliciting the bribe, while all others choose the teacher who does not solicit bribes. The cutoff c^* must be equal to $b + a$.

  Notice that, given that the other students are using this cutoff strategy, a bad student—with $c_j > b + a$—prefers the teacher soliciting the bribe. He/she anticipates that all students who choose this teacher will pay the bribe. This is because $c_j > b + a > b + \frac{1}{N}a$ for all of them. Thus, by choosing the teacher who solicits bribes, the student will receive a payoff of $1 - b - a$ that exceeds the payoff obtained from choosing the other teacher if $1 - b - a > 1 - c_j$, or when $c_j > b + a$. Moreover, bad students prefer to pay the bribe over not going to school because of the assumption that $\pi_{out} < 1 - b - a$. As for the other students, the ones with $c_j \leq b + a$, they anticipate that all the students who choose the teacher soliciting bribes will pay the bribe. So, they prefer to choose the teacher that does not solicit bribes because they obtain a payoff of $1 - c_j$, which is larger than both $1 - b - a$ and $1 - c_j - \frac{n-1}{n}a$.

  It is not possible that, in equilibrium, students use a c^* that is larger than $b + a$. In that case, students with an effort cost $b + a < c_j < c^*$ would prefer to deviate. They would choose the teacher who solicits bribes and pay the bribe themselves. By doing so, they would earn $1 - b - a > 1 - c_j$ for $b + a < c_j < c^*$.

Likewise, it cannot happen that, in equilibrium, all students use a c^* that is smaller than $b + a$. To see this, let $c^* < a + b$. Then, there is a $k \in \{2, \ldots, N\}$ such that $b + \frac{1}{k}a \leq c^* < b + \frac{1}{k-1}a$. Consider the best type who chooses the teacher who solicits bribes ($c_j = c^*$). This student chooses to bribe when the school's size is k or larger, but not when it is smaller than k. For the school sizes where the student does not pay the bribe, he/she would have been better off by choosing the other teacher. This is because, for these school sizes, the probability that one of the other students pays the bribe is greater than 0. For the school sizes where the student pays the bribe, he/she will be in a school where everyone pays the bribe (because the others have at least as high effort costs as the respective student does). Therefore, he/she will earn $1 - b - a$, which is less than the payoff for choosing the other teacher. As a consequence, for any $b + \frac{1}{N}a \leq c^* < b + a$, the best type which is supposed to choose the teacher soliciting bribes prefers to deviate. Notice that, $c^* < b + \frac{1}{N}a$ can also not happen in equilibrium because, for any school size, students with $c_j < b + \frac{1}{N}a$ prefer the teacher who does not solicit bribes.

- **Both teachers solicit bribes.** In this case, students have two options: (i) to go to school and then decide whether to pay the bribe; or (ii) not to go to school and get a payoff of π_{out}. By going to school and paying the bribe, students can obtain at least $1 - b - a$ which is larger than π_{out} by assumption (1). Therefore, they all prefer to go to school. In this case, we assume that students are randomly split between the two teachers.

Stage 1: Teachers' choices. Now, we consider the longest sub-game in which both teachers simultaneously decide whether to solicit bribes. For the teachers' decisions, the institutional setting needs to be taken into account, i.e., *fixed-wage* or *piece-rate.*

Teachers are paid *piece-rate.* The teachers' payoffs are given by the number of students in the school times s and, if they solicit bribes, by the sum of the bribes collected.

- If both teachers do not solicit bribes, they have an expected attendance of $\frac{N}{2}\mathcal{F}(1 - \pi_{out})$ students each, and their expected payoff is $s\frac{N}{2}\mathcal{F}(1 - \pi_{out})$.

- If one teacher solicits bribes, and the other teacher does not, the former has an expected attendance of $N(1 - \mathcal{F}(a + b))$ students, and the latter has an expected attendance of $N\mathcal{F}(a+b)$ students. Given that students in the class where the teacher solicits bribes are paying the bribe b, the expected payoffs are $(s+b)N(1-\mathcal{F}(a+b))$ for the teacher who solicits bribes, and $sN\mathcal{F}(a + b)$ for the teacher who does not solicit bribes.

- If both teachers solicit bribes, students are randomly split between teachers. So, a teacher has either $n = \lceil \frac{N}{2} \rceil$ or $n = \lfloor \frac{N}{2} \rfloor$ students in his/her class. Note that, conditional on having n students in the class, there is a probability $P_n = 1 - \mathcal{F}\left(\frac{a}{n} + b\right)$ that a student is paying the bribe. Hence, conditional on n, the expected number of students paying the bribe is $E_{bn} = nP_n$. Moreover, given that students are randomly split between teachers, the unconditional expected number of students paying the bribe is

$$E_b = \frac{1}{2}\left\lceil \frac{N}{2} \right\rceil P_{\lceil \frac{N}{2} \rceil} + \frac{1}{2}\left\lfloor \frac{N}{2} \right\rfloor P_{\lfloor \frac{N}{2} \rfloor}$$

Therefore, two teachers' payoffs are $s\frac{N}{2} + bE_b$.

The following payoff matrix summarizes the expected payoffs of T_1 at stage 1 (note that the game is symmetric).

		T_2: $B_2 = 1$	T_2: $B_2 = 0$
T_1	$B_1 = 1$	$s\frac{N}{2} + bE_b$	$(s+b)N(1-\mathcal{F}(a+b))$
	$B_1 = 0$	$sN\mathcal{F}(a+b)$	$s\frac{N}{2}\mathcal{F}(1-\pi_{out})$

The payoff matrix highlights the 2 inequalities governing the incentives of the teachers. When T_2 solicits bribes ($B_2 = 1$), T_1 prefers not to solicit bribes ($B_1 = 0$) if

$$sN\left(\mathcal{F}(a+b) - \frac{1}{2}\right) > bE_b \tag{1}$$

is satisfied; and when T_2 does not solicit bribes ($B_2 = 0$), T_1 prefers not to solicit ($B_1 = 0$) if

$$sN\left(\frac{\mathcal{F}(1-\pi_{out})}{2} - 1 + \mathcal{F}(a+b)\right) > bN(1-\mathcal{F}(a+b)) \tag{2}$$

is satisfied.

Now we discuss the cases for different levels of $\mathcal{F}(a+b)$:

- Let's start with $\mathcal{F}(a+b) \leq \frac{1}{2}$. This directly implies that both condition (1) and (2) are not satisfied. Therefore, when the externality is small compared to the distribution of effort costs, there is no piece-rate that can prevent teachers soliciting bribes and in equilibrium is both teachers solicit bribes.
- Let's consider the case where $\mathcal{F}(a+b) > \frac{1}{2}$ and $\mathcal{F}(a+b) \leq 1 - \frac{\mathcal{F}(1-\pi_{out})}{2}$. In this case condition (2) is not satisfied and condition (1) is satisfied iff $s > \frac{E_b}{N\left(\mathcal{F}(a+b)-\frac{1}{2}\right)}b$. This implies that when $s \leq \frac{E_b}{N\left(\mathcal{F}(a+b)-\frac{1}{2}\right)}b$ both teachers solicit bribes, and when $s > \frac{E_b}{N\left(\mathcal{F}(a+b)-\frac{1}{2}\right)}b$ the two teachers specialize, with one teacher soliciting and the other not soliciting. In this latter case the game resembles the "Battle of the sexes".
- Let's finally consider the case where $\mathcal{F}(a+b) > \frac{1}{2}$ and $\mathcal{F}(a+b) > 1 - \frac{\mathcal{F}(1-\pi_{out})}{2}$. With these vales of $\mathcal{F}(a+b)$ we have 4 possibilities:
  - If $s \leq \min\left(\frac{E_b}{N\left(\mathcal{F}(a+b)-\frac{1}{2}\right)}, \frac{2(1-\mathcal{F}(a+b))}{\mathcal{F}(1-\pi_{out})+2\mathcal{F}(a+b)-2}\right)b$ then both teachers solicit bribes in equilibrium;
  - If $\frac{E_b}{N\left(\mathcal{F}(a+b)-\frac{1}{2}\right)}b < s \leq \frac{2(1-\mathcal{F}(a+b))}{\mathcal{F}(1-\pi_{out})+2\mathcal{F}(a+b)-2}b$ the two teachers specialize, with one teacher soliciting and the other not soliciting in equilibrium.

– If $\frac{E_b}{N\left(\mathcal{F}(a+b)-\frac{1}{2}\right)}b \geq s > \frac{2(1-\mathcal{F}(a+b))}{\mathcal{F}(1-\pi_{out})+2\mathcal{F}(a+b)-2}b$ then teachers are in a coordination game where either they both solicit or they both do not solicit bribes in equilibrium;

– If $s > \max\left(\frac{E_b}{N\left(\mathcal{F}(a+b)-\frac{1}{2}\right)}, \frac{2(1-\mathcal{F}(a+b))}{\mathcal{F}(1-\pi_{out})+2\mathcal{F}(a+b)-2}\right)b$ then not soliciting bribes is a dominant strategy.

Therefore, sufficient conditions for $B_1 = 0$ to be a dominant strategy are that: (i) the loss in reputation is not negligible compared to the distribution of the effort cost of the students

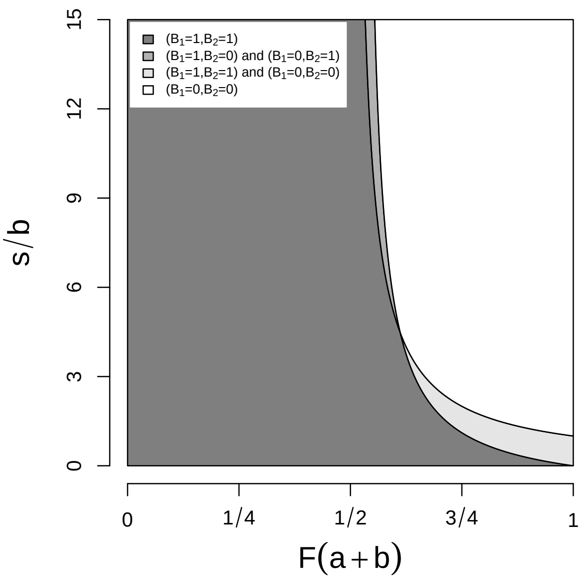
$$\mathcal{F}(a+b) > 1 - \frac{\mathcal{F}(1-\pi_{out})}{2}$$

and (ii) that the piece-rate s is at least t times the size of the bribe, where

$$t = \max\left(\frac{E_b}{N\left(\mathcal{F}(a+b)-\frac{1}{2}\right)}, \frac{2\left(1-\mathcal{F}(a+b)\right)}{\mathcal{F}(1-\pi_{out})+2\mathcal{F}(a+b)-2}\right).$$

Figure A.5 reports an example showing how the set of equilibria changes with s and with $\mathcal{F}(a+b)$. The white region represents the combination of values for which not soliciting bribes is a dominant strategy.

Figure A.5: Equilibria in the *piece-rate* regime as a function of $\mathcal{F}(a+b)$ and s/b (the example assumes $\mathcal{F}(1-\pi_{out}) = 0.95$ and $\frac{E_b}{N} = 0.5$).



Teachers are paid a *fixed-wage*. The teachers' payoffs are a fixed amount F. If a teacher solicits bribes, the teacher has the opportunity to make more money collecting the bribes paid.

This case is formally equivalent to the *piece-rate* case with $s = 0$. In this case, the payoff matrix for teacher T_1 is

		T_2	
		$B_2 = 1$	$B_2 = 0$
T_1	$B_1 = 1$	$F + bE_b$	$F + bN(1 - \mathcal{F}(a + b))$
	$B_1 = 0$	F	F

Inspecting the matrix, it is easy to see that soliciting bribes is a dominant strategy for the teachers when they are paid a *fixed-wage*. $\square$

A.2 Endogenous bribes in the general model

The general model assumes an exogenously given size of the bribe b. Here, we relax this assumption. We assume that each teacher can choose the level of the bribe he/she is willing to solicit. The bribe b_i can take any non-negative value and it is communicated to the students before they make their choices. We show that, even with endogenously chosen bribes, both teachers not soliciting bribes remains an equilibrium in the piece-rate regime. But this implies slightly stronger conditions for $F(a+b)$ and s.

In this section, we make a different set of assumptions. We assume that $\pi_{out} \leq 1-a$, which is a slightly weaker version of Assumption 1. This new assumption implies that it is preferable to get a bad diploma for free (i.e., paying $b=0$) than skipping school altogether. We further assume that $b \leq \bar{c} - \frac{1}{N}a$, which sets an upper bound to the bribe level. This is a new assumption that implies that teachers do not set the bribe to a level that no student is ever willing to pay. Finally, we assume that, when indifferent, students avoid strategic risk: i.e., students prefer to be in a class where bribes are not solicited compared to a class where bribes are solicited but are never paid.

Now, we start from the case where both teachers do not solicit bribes and analyze what happens when T_1 deviates to soliciting a bribe b_1.
[**Case large bribes:** $1-a-b_1 \leq \pi_{out}$]. Consider first b_1 such that $1-a-b_1 \leq \pi_{out}$. We show that, for such bribe values, students do not join the bribing class.

1. Suppose that there are students joining T_1 and that all the students who join prefer to pay the bribe for all sizes of the class. If this is the case, the students believe that only students with $c_j > a+b_1$ join T_1. These students, however, have a profitable deviation because they would prefer to skip school: $1-a-b_1 \leq \pi_{out}$.

2. Suppose that there are students joining T_1 and that some of them pay the bribe for some sizes of the class but not for others. This means that some students with $c_j \leq b_1 + a$ join T_1. If the students joining all have $c_j > b_1 + a$, they all pay the bribe for all sizes of the class and this contradicts the assumption. Moreover, there are no students with $c_j \leq b_1 + \frac{1}{N}a$ that join T_1. These students will never pay the bribe and, given that we assume that there are some students bribing for class sizes, they have an incentive to deviate to avoid the externality. Therefore, some students must join the bribing class for some c_j in the interval $\left(b_1 + \frac{1}{N}a, b_1 + a\right]$.

   Consider the partition of the interval in sub-intervals of the type
   $$I(k) = \left(b + \frac{1}{k+1}a, b + \frac{1}{k}a\right]$$
   with $k \in \{1, 2, \ldots, N-1\}$. Take k^* such that $I(k^*)$ contains a c_j for which students join the bribing class and there are no c_j for which students join the bribing class in all the sub-intervals $I(k)$ where $k > k^*$. Let c^* be a c_j in $I(k^*)$ for which students join the bribing class. Note that, a student with c^* pays the bribe if the class is strictly bigger than k^*. Moreover, whenever the student pays the bribe, all the other students in the class pay the bribe as well. Therefore the student's payoff is $1-a-b_1$ when paying the bribe, and, at best, $1-c^*$ when not paying the bribe. The fact that $c^* < b_1 + a$ implies that a student with such a c_j has a profitable deviation to join the teacher that does not solicit bribes.

3. Therefore, it must be that all the students joining T_1 prefer not to pay the bribe.

If this is the case, the students believe that only students with $c_j \leq 1 - \pi_{out}$ and $c_j \leq b + \frac{1}{N}a$ join the class. Otherwise, they would have a profitable deviation. These students, however, are equally well-off when joining the T_1, and when joining T_2. In this case, we assume that they do not join T_1 in order to avoid strategic risk.

[**Case small bribes:** $1 - a - b_1 > \pi_{out}$]. If b_1 is such that $1 - a - b_1 > \pi_{out}$, reflects the same case as in the general model with an exogenous bribe. When both teachers do not solicit bribes, their payoff is

$$s\frac{N}{2}\mathcal{F}(1 - \pi_{out}).$$

When T_1 deviates to soliciting bribes, his/her payoff is

$$(s + b_1)N(1 - \mathcal{F}(a + b_1)).$$

Therefore, T_1 does not to have a profitable deviation when

$$s\frac{N}{2}\mathcal{F}(1 - \pi_{out}) \geq (s + b_1)N(1 - \mathcal{F}(a + b_1))$$

for all $b_1 \in [0, 1 - \pi_{out} - a)$. Note that: (i) $F(a + b)$ attains its minimum at $F(a)$, (ii) $\mathcal{F}(1 - \pi_{out}) \geq \mathcal{F}(a)$, and (iii) $b_1 < 1 - a - \pi_{out}$. Therefore, sufficient conditions to prevent profitable deviations are that:

- the probability to have bad students is below $\frac{1}{3}$, i.e.,

$$\mathcal{F}(a) > \frac{2}{3}$$

  and that

- the piece-rate s is big enough, i.e.,

$$s > \frac{2(1 - \mathcal{F}(a))}{3\mathcal{F}(a) - 2}(1 - a - \pi_{out}) \geq \frac{2(1 - \mathcal{F}(a))}{3\mathcal{F}(a) - 2}b$$

Note that, these conditions are slightly stronger versions of the ones derived for the case of a fixed bribe b. Both these conditions are satisfied when using the experimental parameters, i.e., $s = 50$, $a = 100$, $\pi_{out} = 115$, $\mathcal{F}(a) = \frac{5}{6}$. Moreover, with these parameters, no equilibria exist where both teachers solicit bribes. This requires long and tedious calculations that are not included in the supplementary materials but are available from the authors upon request.

A.3 More than two schools in the general model

In this section, we analyze the case where $K > 2$ schools compete for students. Given the assumptions made in the general model, a bribe-free equilibrium in the piece-rate regime with two schools exists whenever $s\frac{N}{2}\mathcal{F}(1-\pi_{out}) > (s+b)N(1-\mathcal{F}(a+b))$ and $\mathcal{F}(a+b) > 1 - \frac{\mathcal{F}(1-\pi_{out})}{2}$. This implies that

$$\frac{s}{(s+b)}\frac{\mathcal{F}(1-\pi_{out})}{1-\mathcal{F}(a+b)} > 2.$$

Suppose now that there are $K > 2$ schools. Note that, if K is such that

$$\frac{s}{(s+b)}\frac{\mathcal{F}(1-\pi_{out})}{1-\mathcal{F}(a+b)} \geq K > 2,$$

the strategy profile where all the teachers are not soliciting bribes is still an equilibrium because $s\frac{N}{K}\mathcal{F}(1-\pi_{out}) \geq (s+b)N(1-\mathcal{F}(a+b))$.

If, instead, K is such that

$$K > \frac{s}{(s+b)}\frac{\mathcal{F}(1-\pi_{out})}{1-\mathcal{F}(a+b)} > 2,$$

we prove that there is an equilibrium where some solicit bribes and some do not. Consider the partition of the interval from 0 to K in intervals of the type

$$I(Q) = \left[\frac{K-Q}{Q+1}, \frac{K-Q+1}{Q}\right)$$

where $Q \in \{1, 2, \ldots, K\}$. Take Q^* such that $\frac{s}{(s+b)}\frac{\mathcal{F}(1-\pi_{out})}{1-\mathcal{F}(a+b)} \in I(Q^*)$ and note that $Q^* < K$ because of $K > \frac{s}{(s+b)}\frac{\mathcal{F}(1-\pi_{out})}{1-\mathcal{F}(a+b)} > 2$ and $I(K) = \left[0, \frac{1}{K}\right)$.

Suppose that Q^* teachers solicit bribes and $K - Q^*$ teachers do not solicit bribes. We can show that this is an equilibrium by looking at teachers profitable deviations. Note that: (i) $\frac{K-Q^*}{Q^*+1} \leq \frac{s}{(s+b)}\frac{\mathcal{F}(1-\pi_{out})}{1-\mathcal{F}(a+b)}$ implies that $s\frac{N}{K-Q^*}\mathcal{F}(1-\pi_{out}) \geq (s+b)\frac{N}{Q^*+1}(1-\mathcal{F}(a+b))$, which means that the teachers who do not solicit bribes have no incentive to deviate; and (ii) $\frac{K-Q^*+1}{Q^*} > \frac{s}{(s+b)}\frac{\mathcal{F}(1-\pi_{out})}{1-\mathcal{F}(a+b)}$ implies $s\frac{N}{K-Q^*+1}\mathcal{F}(1-\pi_{out}) < \frac{N}{Q^*}(s+b)(1-\mathcal{F}(a+b))$, which means that also the teachers who solicit bribes have no incentive to deviate.

Finally, note that for $K > 2$ there is no equilibrium where all the teachers solicit bribes. Indeed, our assumptions assure that $sN\mathcal{F}(a+b) > s\frac{N}{2} + bE_b$ and, with K schools, the payoff for teachers who solicit bribes is smaller than $s\frac{N}{K} + bE_b < s\frac{N}{2} + bE_b$.

A.4 Rebates in the general model

In this section, we examine the consequences of teachers attracting students to their classes by providing monetary incentives in the form of rebates. We can show that no equilibria exist where both teachers do not solicit bribes if, as we assume, there are students that prefer to obtain a diploma with a very bad reputation rather than exerting effort, i.e., $1 - \bar{c} < 1 - a - b$. Moreover, we can show that, depending on the models' parameters, there may exist a Nash equilibrium where one teacher solicits and the other teacher does not solicit bribes.

A.4.1 Both teachers do not solicit bribes

Suppose T_1 and T_2 do not solicit bribes and pay a rebate R_1 and R_2 to students when they join the class. Assume $R_1 > R_2$ and note that, if $R_1 > s$, then all the students with $c_j \leq 1 - \pi_{out} + R_1$ choose teacher T_1 and the other students choose the outside option. In this case, T_1 makes a negative payoff, hence, $R_1 = s$ is a profitable deviation for T_1. If $R_1 \leq s$, then all the students with $c_j \leq 1 - \pi_{out} + R_1$ choose teacher T_1 and the other students choose the outside option. In this case, T_2 does not attract students and he/she can profitably deviate to $R_2 = R_1$, where he/she obtains a payoff of $(s - R_2)\frac{N}{2}\mathcal{F}(1 - \pi_{out} + R_2)$.

Assume now that $R_2 = R_1$. If $R_1 = R_2 > s$, then T_1 makes a negative profit and he is better off deviating to $R_1 = s$. If $R_1 = R_2 < s$, the teachers share the students with an effort cost $c_j \leq 1 - \pi_{out} + R_1$. In this case, T_1 has a profitable deviation by setting $R_1 = R_1 + \epsilon < s$ because, by doing so, T_1 obtains a payoff $(s - R_1 - \epsilon)N\mathcal{F}(1 - \pi_{out} + R_1 + \epsilon) > (s - R_1)\frac{N}{2}\mathcal{F}(1 - \pi_{out} + R_1)$. If $R_1 = R_2 = s$, the two teachers share the students with $c_j \leq 1 - \pi_{out} + s$ and make no profits. In this case, a teacher can deviate by soliciting bribes and offering a rebate $R = s$. This teacher will have a positive expected payoff because he/she will attract the students with $c_j \in (b + a, \bar{c}]$ and obtain $bN(1 - \mathcal{F}(b + a))$.

A.4.2 One teacher solicits bribes and the other teacher does not solicit bribes.

Suppose that T_1 solicits bribes and T_2 does not. Suppose also that T_1 offers a rebate R_1 and T_2 offers a rebate R_2. We now look for Nash equilibria where: (i) the students with an effort cost higher than c^* choose T_1 and pay the bribe for all sizes of the class; and (ii) the students with an effort cost lower or equal than c^* choose T_2.

Let us first look at the students. For students with $c_j > c^*$, it must be the case that

$$1 - b - a + R_1 \geq \max\left(1 - c_j + R_2, 1 - c_j + R_1\right).$$

For students with $c_j \leq c^*$, instead, it must be the case that

$$1 - c_j + R_2 \geq \sum_{n=0}^{N-1} \max\left(1 - c_j - \frac{n}{n+1}a + R_1, 1 - b - a + R_1\right) P(n|c^*)$$

where $P(n|c^*)$ denotes the probability that n students have an effort cost higher than c^*. Note that, from the first inequality we obtain $c_j \geq a + b + R_2 - R_1$ and, from the second, we obtain $c_j \leq a + b + R_2 + R_1$. This permits to identify a threshold $c^* = a + b + R_2 - R_1$. Furthermore, it must be that $R_1 \leq R_2$ to avoid profitable deviations for students with

$c_j = c^*$.

If students behave according to the equilibrium conjecture, the payoff of T_1 is

$$\pi_1(R_1) = (s + b - R_1)N(1 - \mathcal{F}(a + b + R_2 - R_1))$$

and the payoff of T_2 is

$$\pi_2(R_2) = (s - R_2)N\mathcal{F}(a + b + R_2 - R_1)$$

We can then look at the FOC for a teacher's optimal rebate and obtain the best response functions $R_1^*(R_2)$ and $R_2^*(R_1)$. The best response of T_1 is implicitly defined by the condition

$$(s + b - R_1) = \frac{(1 - \mathcal{F}(a + b + R_2 - R_1))}{f(a + b + R_2 - R_1)} \tag{3}$$

and the best response of T_2 by the condition

$$(s - R_2) = \frac{\mathcal{F}(a + b + R_2 - R_1)}{f(a + b + R_2 - R_1)} \tag{4}$$

Therefore, if there are R_1^* and R_2^* such that $R_1^* < R_2^*$ and both equation 3 and equation 4 are satisfied, then a Nash equilibrium where one teacher solicits bribes and the other does not, can be sustained by the belief that students do not join the class of a teacher who switches from soliciting to not soliciting bribes (and vice-versa).

Finally, we provide a simple example showing that there are conditions under which such an equilibrium exists. Assume that the student's effort cost is uniformly distributed between 0 and $\bar{c}$, then the payoff function for both teachers is a quadratic function in the rebate with support $[0, s + b]$ for T_1 and $[0, s]$ for T_2. Therefore, the teachers' maximum payoff is obtained for

$$R_1^*(R_2) = \begin{cases} 0 & \text{if} \quad R_2 < \bar{c} - a - 2b - s \\ \frac{1}{2}(R_2 + a + 2b + s - \bar{c}) & \text{if} \quad \bar{c} - a - 2b - s \leq R_2 \leq \bar{c} - a + s \\ s + b & \text{if} \quad R_2 > \bar{c} - a + s \end{cases}$$

$$R_2^*(R_1) = \begin{cases} 0 & \text{if} \quad R_1 < a + b - s \\ \frac{1}{2}(R_1 - a - b + s) & \text{if} \quad a + b - s \leq R_1 \leq a + b + s \\ s & \text{if} \quad R_1 > a + b + s \end{cases}$$

Note that, for a relatively high piece-rate s, there is a feasible solution where teachers offer the positive rebates $R_1^* = s + \frac{1}{3}(a + 3b - 2\bar{c})$ and $R_2^* = s - \frac{1}{3}(a + \bar{c})$. Note also that these rebates satisfy the students' constraints. The following set of parameters $a = 0.2$, $b = 0.04$, $s = 0.3$, and $\bar{c} = 0.55$, for instance, give $R_1^* = 0.04$ and $R_2^* = 0.05$. These rebates satisfy the students' constraint $R_1 \leq R_2$.

A.5 Effect of schools' size cap with the experimental parameters

In this section, we discuss the effect of a cap for the school sizes to k students for the parameter constellation used in the experiment. Additionally, we assume that the total capacities of the two schools are such that all the N students can be accommodated when they decide to go to school. This means that the cap k is $k \geq \frac{N}{2}$.

Under these assumptions, the students' optimal behavior conditional on the teachers' choices is as follows:

- When both teachers solicit bribes ($B_1 = 1, B_2 = 1$), all the students decide to go to school, and the students that decide to go to school are shared between the teachers. In this case, teachers have $\frac{N}{2}$ students each.
- When both teachers do not solicit bribes ($B_1 = 0, B_2 = 0$), the intermediate and good students decide to go to school, while the bad students decide not to go to school. The students who decide to go to school are split off between the teachers. If n_B is the number of bad students, one teacher gets $\lceil \frac{N-n_B}{2} \rceil$ students and the other teacher gets $\lfloor \frac{N-n_B}{2} \rfloor$ students.

  Note that, in both the previous cases the cap on the school size is not binding. hence, the payoffs of the teachers remain the same as in the non-capped case discussed in the main text.
- When one teacher solicits bribes and the other does not, i.e. either ($B_1 = 1, B_2 = 0$) or ($B_1 = 0, B_2 = 1$), the students apply to one of the schools and, if there are too many students applying to one school, a random selection decides which ones are the students in excess. These students can either decide to go to the other school or not to go to school altogether. Note that, with this application procedure, the following is an equilibrium: (i) the good and intermediate students apply to the bribe-free school and, if rejected, they go to the bribing school; (ii) the bad students apply to the bribing school and, if rejected, they decide not to go to school.

  Therefore, assuming that $B_1 = 0$ and $B_2 = 1$ and letting n_B be the number of bad students and n_G be the number of good students, the payoff of T_1 conditional on n_G and n_B is

$$\pi_1(B_1 = 0, B_2 = 1 | n_B, n_G) = \begin{cases} 50 \cdot (N - n_B) & \text{if } N - n_B \leq k \\ 50 \cdot k & \text{if } N - n_B > k \end{cases}$$

  Assuming, instead, that $B_1 = 0$ and $B_2 = 1$, the payoff of T_1 conditional on n_G and n_B is

$$\pi_1(B_1 = 1, B_2 = 0 | n_B, n_G) = \begin{cases} 50 \cdot n_B + 10 \cdot n_B & \text{if } n_B \leq k \text{ and } N - n_B \leq k \\ 50 \cdot (N - k) + 10 \cdot (n_B + E_b) & \text{if } n_B \leq k \text{ and } N - n_B > k \\ 50 \cdot k + 10 \cdot k & \text{if } n_B > k \end{cases}$$

  where E_b denotes the expected number of intermediate students rejected from T_2's class conditional on having $N - n_B - n_G$ intermediate and n_G good students in the population. With the assumed application procedure, E_b is the mean of a Hypergeometric distribution where $N - n_B - k$ balls are drawn without replacement from an urn containing $N - n_B - n_G$ winning balls and n_G losing balls. That is $E_b = (N - n_B - k)\frac{N-n_B-n_G}{N-n_B}$.

Given these considerations, we can compute the expected payoff of T_1 for all possible cases:

$$E\pi_1(B_1 = 1, B_2 = 1) = 50 \cdot \frac{N}{2} + 10 \cdot \frac{N}{2} \cdot \frac{5}{6}$$

$$E\pi_1(B_1 = 0, B_2 = 1) = \sum_{n_G=0}^{N} \sum_{n_B=0}^{N-n_G} \frac{N!}{n_G! n_B! (N - n_G - n_B)!} \left(\frac{1}{6}\right)^{n_G+n_B} \left(\frac{4}{6}\right)^{N-n_G-n_B} \pi_1(B_1 = 0, B_2 = 1 | n_B, n_G)$$

$$E\pi_1(B_1 = 1, B_2 = 0) = \sum_{n_G=0}^{N} \sum_{n_B=0}^{N-n_G} \frac{N!}{n_G! n_B! (N - n_G - n_B)!} \left(\frac{1}{6}\right)^{n_G+n_B} \left(\frac{4}{6}\right)^{N-n_G-n_B} \pi_1(B_1 = 1, B_2 = 0 | n_B, n_G)$$

$$E\Pi_1(B_1 = 0, B_2 = 0) = 50 \cdot \frac{N}{2} \cdot \frac{5}{6}$$

Table A.2 summarizes the effect of the school size cap of k when $N = 8$. When the cap is relatively large, teachers under piece-rate have an incentive not to solicit bribes in order to attract students. When instead the cap is small, teachers cannot attract many students and this reduces the incentives not to solicit bribes. Therefore, schools' capacity constraints can change the equilibrium in the game by undermining the incentives to not solicit bribes under the piece-rate regime.

Table A.2: Equilibria when there is a cap $k \geq \frac{N}{2}$ on the size of the class and the total number of students in $N = 8$.

Cap k	$(B_1 = 1, B_2 = 1)$	$(B_1 = 0, B_2 = 0)$	$(B_1 = 1, B_2 = 0)$ and $(B_1 = 0, B_2 = 1)$
8		X	
7		X	
6		X	
5			X
4	X		

B. Informed consent and instructions [ONLINE APPENDIX]

Informed consent

Welcome,
You are invited to participate in a study investigating the processes that influence people's decision making. If you agree to participate, we will ask you to complete an interactive computerized task. You will be also asked to answer a short questionnaire at the end of the study. The estimated duration of the entire experiment is 120 minutes.

Before agreeing to participate in this study, it is important that you read and understand the following explanations, so you can make an informed decision about taking part in this study.

Purpose: This study is designed to investigate the processes involved in decision making.

Confidentiality: Data collected will remain strictly confidential. All data will be used for research purposes and to write a scientific paper about the nature of decision processes. Only researchers who are associated with the study will see your responses.

Your responses will not be associated with your name; instead, your name will be converted to a code number when the researchers store the data. No names or identifying information will be used in any publication or presentation.

Potential Risks and Discomforts: There are no anticipated risks associated with participation in this study.

Anticipated Benefits: The benefits associated with participating in this study are:

(1) you receive a participation payment of 12.000 COP,

(2) a payment which is based on your task performance.

(3) the satisfaction to contribute to the scientific understanding of how people make decisions. Upon the completion of the study, you will be given a thorough explanation of the study. You can also opt to receive a manuscript of any manuscript based on the research (or summaries of our results) upon completion.

Participation and Withdrawal: Your participation in this research is entirely voluntary. If you choose not to participate, it will not affect your relationship with any of the researchers involved or their institutes. If you decide to participate, you are free to withdraw your consent and discontinue your involvement at any time without penalty.

Questions: The experimenter will answer any questions about the research either now or during the course of the experiment. You can signal that you have a question by raising your hand and the experimenter will come to you promptly.

If you have other questions or concerns, you can address them to any of the following:
Nils Köbis (n.c.kobis@gmail.com)

Instructions: The instructions below inform you about the general procedure of the study you are about to take part in. It is conducted by the University of Amsterdam and University of Zürich.

Consent: I have had the opportunity to discuss this study and my questions have been answered to my satisfaction. I consent to take part in the study with the understanding that I may withdraw at any time. I am aware that an explanation about the rationale and predictions underlying this experiment will be presented upon completion of the study. I freely consent to take part in this study.

________________________________ Signature, date

General instructions

Thank you for agreeing to participate. During the study, we require your complete, undistracted attention. Please read the following instructions carefully. If you have questions at any point or do not understand the instructions, please raise your hand and one of the assistants will come and help you.

The study has two parts. Both parts have 15 rounds. After the information about the payment of the study you receive the instructions for the first part. Instructions for the second part will be distributed when the first part is over.

In both parts, there are several rounds of decision-making. Your decisions and those of other participants will determine your earnings. You will receive 7 Euros as a participation fee for this study and in addition you will be paid for six extra rounds of decision making. The computer will randomly draw three rounds of the first part and three rounds of the second part. The results of these rounds will be paid out privately to you and the others in cash at the end of today's session.

All the payoffs in the study will be expressed in points. At the end of the study your earnings will be converted in Euro at the conversion rate of

100 point = 3.000 COP.

All interactions among you and other participants will take place through computers. You are **not allowed to speak** to the other participants. If you do not follow that rule you can be excluded from the study. You will not know which specific participant made which decision and the other participants will not know the decisions you made. Your decisions and the decisions of all other participants are completely private.

In the following, the procedure for the first part of the study is described in detail.

Instructions for Part One

Roles

The computer will randomly choose 10 participants and make them into one group. Within such a group, the computer will randomly assign different roles to the participants.

The computer will randomly assign the role of **teacher** to TWO participants and the role of **student** to the remaining EIGHT participants. Other participants do not know your role and you do not know the roles of the other participants. Importantly, each participant will keep the role assigned to them by the computer throughout the entire study.

Basic Structure

In each round, teachers are paid a salary of **40 points** to teach a class and they can decide to ask students to pay a motivation fee of **10 points**. Students on the other hand have the opportunity to either join one of the two classes and obtain a diploma worth **250 points** or decide not to go to school and obtain **115 points**. If students decide to join one of the classes, they have two different ways to obtain the diploma depending on the decision of the teacher. If the teacher in their class is not asking for a motivation fee, they can only get the diploma exerting effort and paying the corresponding **effort cost.** The effort cost can change from student to student and from one round to the other.

If instead the teacher in their class is asking for a motivation fee, they can decide to obtain the diploma either by exerting effort or paying the motivation fee to the teacher. Paying the motivation fee, however, reduces the value of the diploma for all the students in the class. The sequences of decisions and the calculation of the payoffs are described in more detail below.

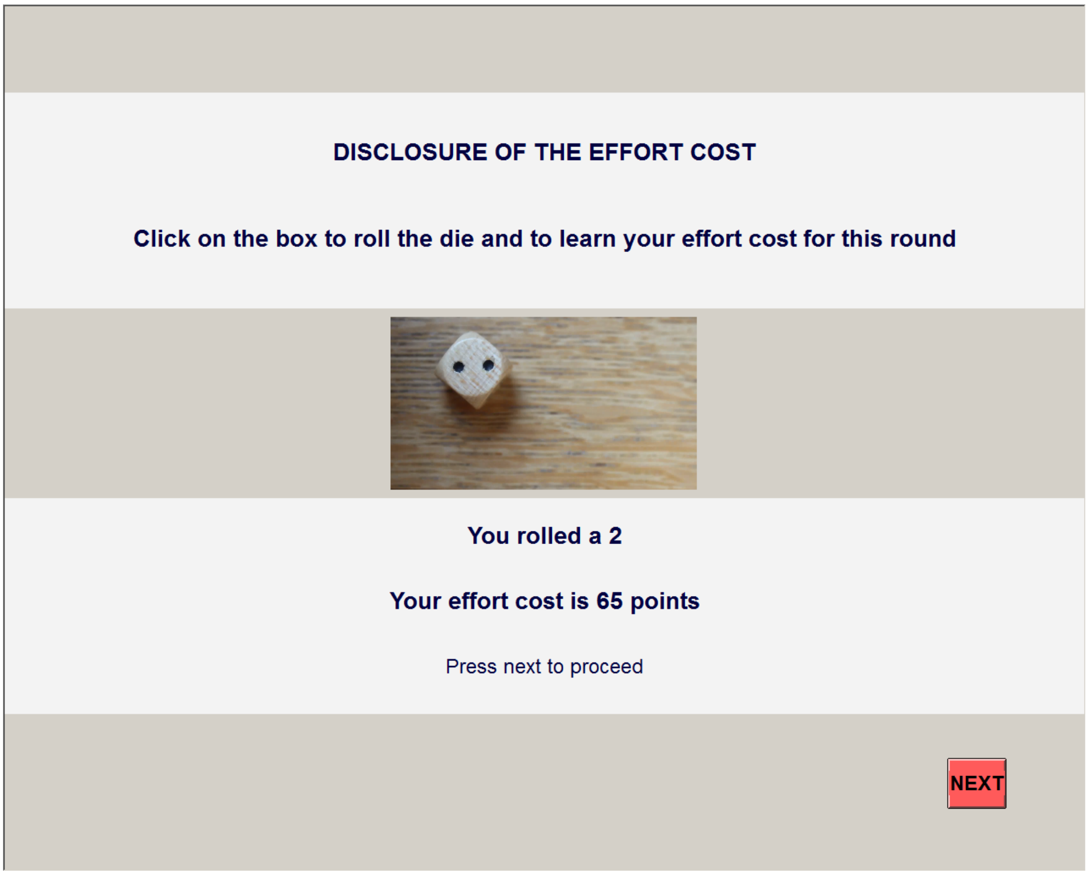
Sequence of Events

At the beginning of each round students are privately informed about their effort cost.

Effort cost

Students can obtain the degree by exerting effort. The cost of effort differs for each student, and per round. There are three levels of effort cost: **10, 65 and 160 points**. At the start of each round, the computer will determine the effort cost of a student with an independent roll of a die. A roll of a 1 leads to an effort cost of 10 points; a roll of 2, 3,4, 5 leads to an effort cost of 65 points; and a roll of 6 leads to an effort cost of 160 points. Each student is informed of her or his own effort cost, but not of the effort costs of the other students. The teachers are also not informed of the effort costs of the students.

The effort cost determines how much a student must pay to obtain the degree without motivation fee.

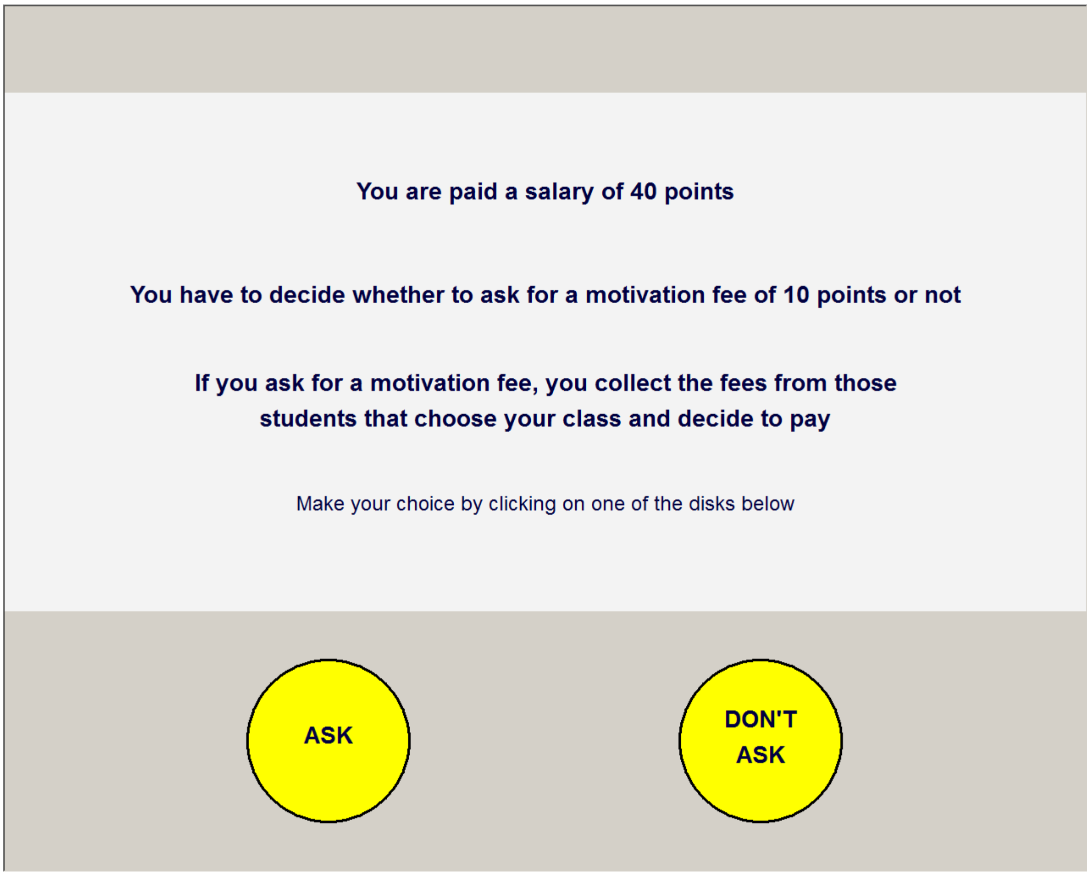


Decisions

Teachers' decision

After students are informed about their effort cost, each teacher decides at the same time whether to

1. ASK for a motivation fee OR
2. NOT ASK for a motivation fee



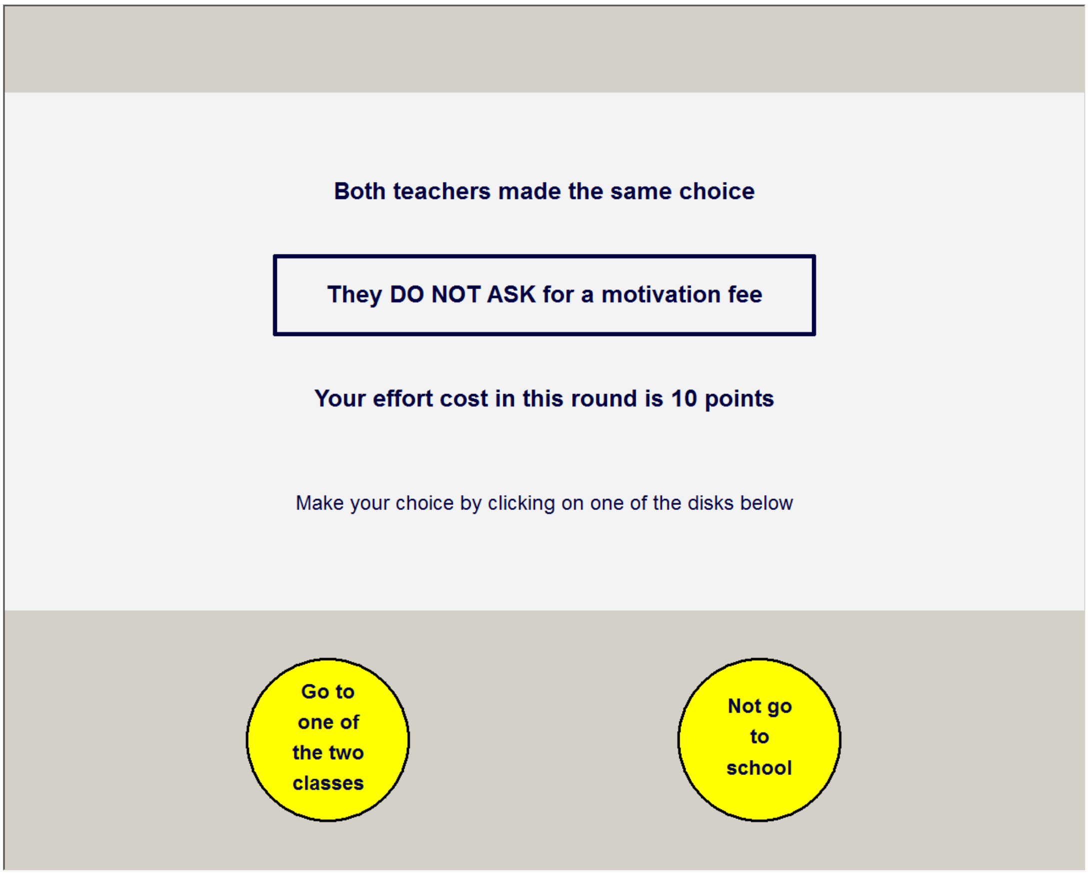
Students' decision

After both teachers decided whether to ask or not to ask for a motivation fee, students are informed about the teachers' decisions and are asked to make their own decisions. Depending on the decisions of the teachers, there are three possible scenarios.

1. Both teachers DO NOT ASK for a motivation fee

In this case students can choose either to

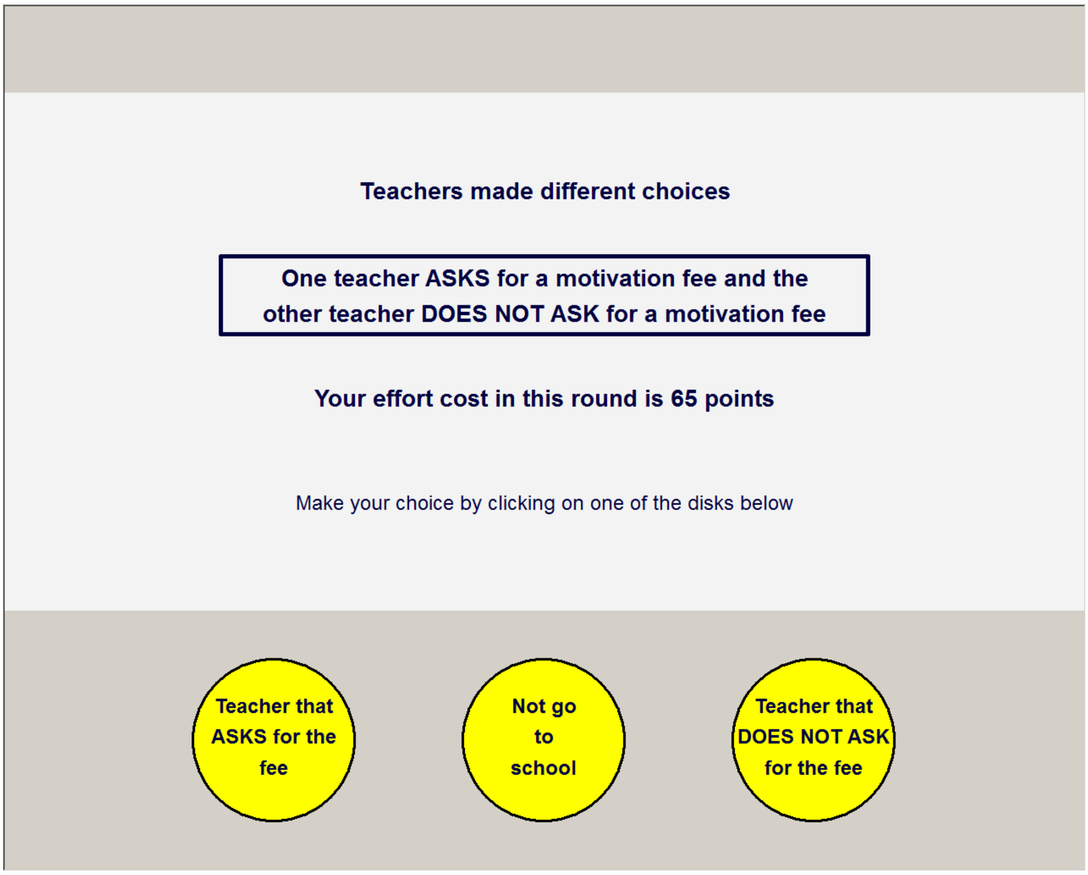
1. not go to school OR
2. be assigned to one of the two teachers and obtain the diploma by exerting effort. The computer will randomly assign half of the students choosing this option to one teacher and the other half to the other teacher.

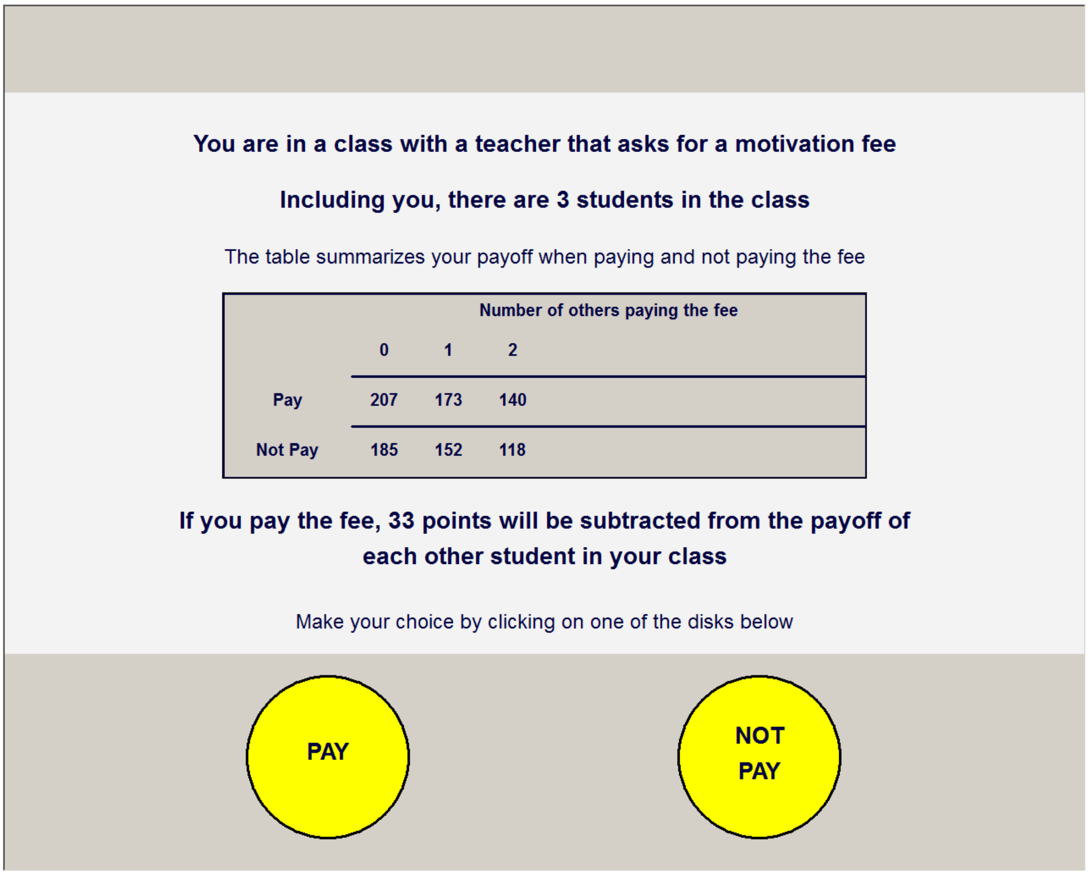


2. One teacher ASKS for a motivation fee and the other teacher DOES NOT ASK for a motivation fee

In this case, the students can choose either to:

1. not go to school OR
2. go to the teacher that DOES NOT ASK for a motivation fee and obtain the diploma by exerting effort. OR
3. go to the teacher that ASKS for a motivation fee and make a second choice whether
   (a) to PAY the motivation fee and obtain the diploma without exerting effort
   (b) NOT TO PAY the motivation fee and obtain the diploma by exerting effort


Teachers made different choices
One teacher ASKS for a motivation fee and the other teacher DOES NOT ASK for a motivation fee
Your effort cost in this round is 65 points
Make your choice by clicking on one of the disks below
Teacher that ASKS for the fee
Not go to school
Teacher that DOES NOT ASK for the fee



You are in a class with a teacher that asks for a motivation fee

Including you, there are 3 students in the class

The table summarizes your payoff when paying and not paying the fee

	Number of others paying the fee		
	0	1	2
Pay	207	173	140
Not Pay	185	152	118

If you pay the fee, 33 points will be subtracted from the payoff of each other student in your class

Make your choice by clicking on one of the disks below

PAY

NOT PAY

3. Both teachers ASK for a motivation fee. In this case students can choose to either:

1. not go to school OR
2. be assigned to one of the two teachers and make a second choice whether
   (a) to PAY the motivation fee and obtain the diploma without exerting effort
   (b) NOT TO PAY the motivation fee and obtain the diploma by exerting effort

   The computer will randomly assign half of the students choosing this option to one teacher and the other half to the other teacher.

Payoffs

Teachers' payoffs:
Teachers receive a fixed-wage for each round of **40 points.** If a teacher asks for a motivation fee, the teacher receives in addition **10 points** for each student who joins the teacher's class and pays the motivation fee.

Students' Payoffs
Not going to school leads to a payoff of **115 points**.
If students go to school the student receives a diploma worth **250 points** but there is a cost to getting the diploma.

If students are in a class where the teacher does not ask for a motivation fee, each student pays his/her individual effort cost to obtain the diploma.

Example: In a class in which a teacher DOES NOT ASK for a motivation fee, a student with an effort cost of **10 points**, has to pay that effort cost to obtain the degree. The payoff for that student in that round therefore is **240 = 250 (value of the degree) - 10 (effort cost)**

If students are in a class where the teacher ASKS for a motivation fee, each student decides whether to pay the motivation fee or not.

The students who do not pay the motivation fee pay the effort cost and the students who pay the motivation fee do not pay the effort cost.

Independent of whether a student pays or does not pay the motivation fee, the value of the diploma is reduced according to the proportion of students in the class who pay the motivation fee. That means, if **all** students in a class pay the motivation fee the value of the diploma for each student in this class is reduced by **100 points**.

If **half** of the students pay the motivation fee the value of the diploma is reduced by $\frac{1}{2} * \mathbf{100 = 50}$ **points**

Example: If there are **four students** in the class and **three of them pay** the motivation fee, the value of the diploma is reduced for all students by $\frac{3}{4} * \mathbf{100 = 75}$ **points** Therefore, if the student that does not pay the motivation fee has an effort cost of **65 points**, he obtains a payoff of **110 = 250 (value of the degree) − 65 (effort cost) − 75**. The students that pay the motivation fee, instead, have a payoff of **165 = 250 (value of the degree) − 10 (motivation fee) − 75** independently of their effort cost.

At the end of each round, students and teachers are informed of the results of the round before the next round is started.

Instructions for Part 1 are over. We will now ask you to answer some questions on your computer screen to ensure that you understand the instructions completely.

Summary of the decisions

Part 1 consists of 15 rounds. At the beginning of each round:

1. **Students** are **informed** about their **effort cost**
2. **Teachers** independently **decide** whether to **ask for a motivation fee** or **not**
3. **Students** are **informed** about the **decision of both teachers**. That means that they know that either: I) **both** teachers **ask for a motivation fee**

   II) **both** teachers **don't ask for a motivation fee**

   in these cases, **students decide** whether to

   (a) **not go to school at all**, OR
   (b) **be randomly assigned** to one of the classes

   III) **one teacher asks for a motivation fee** and the **one teacher doesn't ask for a motivation fee**

   in this case, **students decide** whether to either

   (a) **not go to school at all**, OR
   (b) **go** to the class in where the **teacher asks for a motivation** fee, OR
   (c) **go** to the class in where the teacher **does not ask for a motivation fee**.
4. **Students** who decide to **go to one of the two classes** will **receive information** about **how many other students are with them in the class**.
5. **If students** are in **class** with a **teacher** who **asks for a motivation fee**, **students decide** whether to **pay the motivation fee**.

Instructions for Part Two (FW)

Basic Structure

In the second part of the study, the salary for the teachers is **240 points**. This is the only change in the structure of the study compared to Part one. Below is a short summary of the payoffs for Part Two.

Teachers' payoffs

In part 2, teachers receive a fixed-wage for each round of **240 points**.

If a teacher asks for a motivation fee, the teacher receives in addition **10 points** for each student who pays the motivation fee.

Students' Payoffs

In part 2, the students' payoffs remain the same.

That means, that not going to school leads to a payoff of **115 points**.

Going to school and receiving a diploma is worth **250 points** but there is a cost to getting the diploma.

If students are in a class where the teacher does not ask for a motivation fee, each student pays his/her individual effort cost to obtain the diploma.

If students are in a class where the teacher ASKS for a motivation fee, each student decides whether to pay the motivation fee or not.

The students who do not pay the motivation fee, pay the effort cost and the students who pay the motivation fee, do not pay the effort cost.

Independent of whether a student pays or does not pay the motivation fee, the value of the diploma is reduced according to the proportion of students who pay the motivation fee. That means, if **all** students pay the motivation fee the value of the diploma for student is reduced by **100 points**.

If **half** of the students pay the motivation fee the value of the diploma is reduced by $\frac{1}{2} * \mathbf{100 = 50}$ **points**.

At the end of each round, students and teachers are informed of the results of the round before the next round is started.

Instructions for Part 2 are over. We will now ask you to answer some questions on your computer screen to ensure that you understand the instructions completely.

Instructions for Part Two (PR)

Basic Structure

In the second part of the study, the salary of the teachers depends on the number of students in their class. That means that, on top of the fixed-wage of **40 points**, a teacher receives an amount of **50 points** for each student in the class.

Example: If 4 students are in the teacher's class, the teacher receives $\mathbf{40 + 4 * 50 = 240}$ **points** as a salary for this round.

This is the only change in the structure of the study compared to Part one. Below is a short summary of the payoffs for Part Two

Teachers' payoffs

In each round of Part 2, on top of the fixed-wage of **40 points**, teachers receive an amount of **50 points** for each student in the class.

If a teacher asks for a motivation fee, the teacher receives in addition **10 points** for each student who pays the motivation fee.

Students' Payoffs

In Part 2, the students' payoffs remain the same.

That means, that not going to school leads to a payoff of **115 points**.

Going to school and receiving a diploma is worth **250 points** but there is a cost to getting the diploma.

If students are in a class where the teacher does not ask for a motivation fee, each student pays his/her individual effort cost to obtain the diploma.

If students are in a class where the teacher ASKS for a motivation fee, each student decides whether to pay the motivation fee or not.

The students who do not pay the motivation fee, pay the effort cost and the students who pay the motivation fee, do not pay the effort cost.

Independent of whether a student pays or does not pay the motivation fee, the value of the diploma is reduced according to the proportion of students who pay the motivation fee. That means, if **all** students pay the motivation fee the value of the diploma for student is reduced by **100 points**.

If **half** of the students pay the motivation fee the value of the diploma is reduced by $\frac{1}{2} * \mathbf{100 = 50}$ **points**.

Instructions for Part 2 are over. We will now ask you to answer some questions on your computer screen to ensure that you understand the instructions completely.

C. Inequity aversion with the experimental parameters [ONLINE APPENDIX]

C.1 Inequity averse teachers with the experimental parameters

In this section, we derive predictions for the constellation of parameters used in the experiment, assuming that teachers are inequity averse. To do that, we assume that the teachers have the following utility function Fehr & Schmidt (1999)

$$U(x_i, x_{-i}) = x_i - \beta \sum_{j \neq i} \max(x_i - x_j, 0) - \alpha \sum_{j \neq i} \max(x_j - x_i, 0)$$

where: x_i is the payoff of the agent, x_{-i} are the payoffs of the other agents, α is the parameter measuring aversion to disadvantageous inequality, and β is the parameter measuring aversion to advantageous inequality.

C.1.1 Inequity averse teachers under fixed-wage (*F=40* and *F=240*)

We first derive the utility function of the teachers for three possible cases.

(i) Both teachers solicit bribes. When both T_1 and T_2 solicit bribes, all the students go to school and each teacher has 4 students in his/her class. The good students do not pay the bribe and the bad and intermediate students pay the bribe.

Let n_{Gi} be the number of good students in the class of teacher T_i. The payoff for all the students in this class is $\pi_s(n_{Gi}) = 250 - 10 - \frac{100 \cdot (4 - n_{Gi})}{4}$. The payoff for T_i is $\pi_i(n_{Gi}) = F + 10 \cdot (4 - n_{Gi})$, where F can be either 40 or 240, depending on the experimental part. Therefore, conditional on n_{G1} and n_{G2}, the utility of a teacher, when both teacher solicit bribes, takes the following form

$$\begin{aligned}
U_{(1,1)}(n_{G1}, n_{G2}, \alpha, \beta) = \pi_1(n_{G1}) + & \\
& - \beta \max(\pi_1(n_{G1}) - \pi_2(n_{G2}), 0) + \\
& - \beta \left[4 \max(\pi_1(n_{G1}) - \pi_s(n_{G1}), 0) + 4 \max(\pi_1(n_{G1}) - \pi_s(n_{G2}), 0)\right] + \\
& - \alpha \max(\pi_2(n_{G2}) - \pi_1(n_{G1}), 0) + \\
& - \alpha \left[4 \max(\pi_s(n_{G1}) - \pi_1(n_{G1}), 0) + 4 \max(\pi_s(n_{G2}) - \pi_1(n_{G1}), 0)\right]
\end{aligned}$$

From this, we can obtain the unconditional expected utility by summing over the possible realizations of the students' ability levels. This takes the following form

$$EU_{(1,1)}(\alpha, \beta) = \sum_{n_{G1}=0}^{4} \sum_{G_2=0}^{4} \binom{4}{n_{G1}} \left(\frac{1}{6}\right)^{n_{G1}} \left(\frac{5}{6}\right)^{4-n_{G1}} \binom{4}{n_{G2}} \left(\frac{1}{6}\right)^{n_{G2}} \left(\frac{5}{6}\right)^{4-n_{G2}} U_{(1,1)}(n_{G1}, n_{G2}, \alpha, \beta)$$

that, after simplification, becomes

$$EU_{(1,1)}(\alpha, \beta) = \frac{220}{3} - \frac{799755}{209952}\beta - \frac{140767775}{209952}\alpha$$

for $F = 40$ and

$$EU_{(1,1)}(\alpha, \beta) = \frac{820}{3} - \frac{196754975}{209952}\beta - \frac{799775}{209952}\alpha$$

for $F = 240$.

(ii) One teacher solicits bribes and the other teacher does not solicit bribes. When one teacher solicits bribes and the other does not, the good and intermediate students go to the teacher who does not solicit bribes and obtain a payoff $\pi_G = 250 - 10$, and $\pi_M = 250 - 65$, respectively. The bad students go to the teacher who solicits bribes and pay the bribe, obtaining a payoff of $\pi_B = 250 - 10 - 100$.

Let n_G be the number of good students and n_B be the number of bad students. The teacher soliciting bribes (T_1) obtains a payoff of $\pi_1(n_B) = F + n_B \cdot 10$. The teacher not soliciting bribes (T_2), instead, obtains a payoff of $\pi_2 = F$. Conditional on n_B and n_G, the utility of T_1 takes the following form

$$\begin{aligned} U_{(1,0)}(n_G, n_B, \alpha, \beta) = \pi_1(n_B) + & \\ & - \beta \max(\pi_1(n_B) - \pi_2, 0) + \\ - \beta\,[n_G \max(\pi_1(n_B) - \pi_G, 0) + (8 - n_G - n_B) & \max(\pi_1(n_B) - \pi_M, 0) + n_B \max(\pi_1(n_B) - \pi_B, 0)] + \\ & - \alpha \max(\pi_2 - \pi_1(n_B), 0) + \\ - \alpha\,[n_G \max(\pi_G - \pi_1(n_B), 0) + (8 - n_G - n_B) & \max(\pi_M - \pi_1(n_B), 0) + n_B \max(\pi_B - \pi_1(n_B), 0)] \end{aligned}$$

and the unconditional expected utility is

$$EU_{(1,0)}(\alpha, \beta) = \sum_{n_G=0}^{8} \sum_{n_B=0}^{8-n_G} \frac{8!}{n_G! n_B! (8 - n_G - n_B)!} \left(\frac{1}{6}\right)^{n_G+n_B} \left(\frac{4}{6}\right)^{8-n_G-n_B} U_{(1,0)}(n_G, n_B, \alpha, \beta)$$

After simplifications, this becomes

$$EU_{(1,0)}(\alpha, \beta) = \frac{160}{3} - \frac{40}{3}\beta - \frac{3200}{3}\alpha$$

for $F = 40$ and

$$EU_{(1,0)}(\alpha, \beta) = \frac{760}{3} - \frac{1640}{3}\beta$$

for $F = 240$.

Similarly, one can compute the conditional and unconditional utilities given n_G and n_B for the teacher not soliciting bribes. In this case, the unconditional expected utility, i.e., $EU_{(0,1)}(\alpha, \beta)$, becomes

$$EU_{(0,1)}(\alpha, \beta) = 40 - \frac{3560}{3}\alpha$$

for $F = 40$ and

$$EU_{(0,1)}(\alpha, \beta) = 240 - \frac{1280}{3}\beta - \frac{40}{3}\alpha$$

for $F = 240$.

(iii) Both teachers do not solicit bribes. When both T_1 and T_2 do not solicit bribes, the good and intermediate students go to school and respectively obtain a payoff of $\pi_G = 250 - 10$ and $\pi_M = 250 - 65$, independent of the class they go to. The bad students, instead, do not go to school and obtain $\pi_B = 115$. Teachers obtain a payoff of $\pi_i = F$, independent of the number of students in the class.

Let n_G be the number of good students and n_B be the number of bad students. The utility of a teacher, conditional on n_B and n_G, is given by

$$
\begin{aligned}
U_{(0,0)}(n_G, n_B, \alpha, \beta) = \pi_1 & \\
& - \alpha \left[n_G \max(\pi_1 - \pi_G, 0) + (8 - n_G - n_B) \max(\pi_1 - \pi_M, 0) + n_B \max(\pi_1 - \pi_B, 0) \right] \\
& \qquad - \beta \left[n_G \max(\pi_G - \pi_1, 0) + (8 - n_G - n_B) \max(\pi_M - \pi_1, 0) + n_B \max(\pi_B - \pi_1, 0) \right]
\end{aligned}
$$

and the unconditional expected utility is given by

$$EU_{(0,0)}(\alpha, \beta) = \sum_{n_G=0}^{8} \sum_{n_B=0}^{8-n_G} \frac{8!}{n_G! n_B! (8 - n_G - n_B)!} \left(\frac{1}{6}\right)^{n_G+n_B} \left(\frac{4}{6}\right)^{8-n_G-n_B} U_{(0,0)}(n_G, n_B, \alpha, \beta)$$

After simplifications, the unconditional expected utility becomes

$$EU_{(0,0)}(\alpha, \beta) = 40 - 1140\alpha$$

for $F = 40$ and

$$EU_{(0,0)}(\alpha, \beta) = 240 - 460\beta$$

for $F = 240$.

Payoff matrix for the teachers under fixed-wage.

After having obtained the utility of the teachers for all cases, we can write the payoff matrix for both Low fixed-wage and High fixed-wage.

Low fixed-wage. When $F = 40$, we have the following payoff matrix for the teachers

		T_2: $B_2 = 1$	T_2: $B_2 = 0$
T_1	$B_1 = 1$	$\frac{220}{3} - \frac{799755}{209952}\beta - \frac{140767775}{209952}\alpha$	$\frac{160}{3} - \frac{40}{3}\beta - \frac{3200}{3}\alpha$
	$B_1 = 0$	$40 - \frac{3560}{3}\alpha$	$40 - 1140\alpha$

Note that, conditional on T_2 soliciting bribes, T_1 prefers $B_1 = 1$ over $B_1 = 0$ when

$$\beta \leq \frac{1399680}{159955} + \frac{21675053}{159955}\alpha$$

and, conditional on T_2 not soliciting bribes, T_1 prefers $B_1 = 1$ over $B_1 = 0$ when

$$\beta \leq 1 + \frac{11}{2}\alpha$$

These inequalities identify regions with different equilibria of the game. These are reported in panel (a) of Figure 3 in the main paper.

High fixed-wage. When $F = 240$, we have the following payoff matrix for the teachers

		T_2	
		$B_2 = 1$	$B_2 = 0$
T_1	$B_1 = 1$	$\frac{820}{3} - \frac{196754975}{209952}\beta - \frac{799775}{209952}\alpha$	$\frac{760}{3} - \frac{1640}{3}\beta$
	$B_1 = 0$	$240 - \frac{1280}{3}\beta - \frac{40}{3}\alpha$	$240 - 460\beta$

Note that, conditional on T_2 soliciting bribes, T_1 prefers $B_1 = 1$ over $B_1 = 0$ when

$$\beta \leq \frac{1399680}{21435091} + \frac{399917}{21435091}\alpha$$

and, conditional on T_2 not soliciting bribes, T_1 prefers $B_1 = 1$ over $B_1 = 0$ when

$$\beta \leq \frac{2}{13}$$

These inequalities identify regions with different equilibria of the game. These are reported in panel (b) of Figure 3 in the main paper.

C.1.2 Inequity averse teachers under piece-rate

We first derive the utility functions for the teachers for all three possible cases. Note that, the payoffs for the students remain the same as in the fixed-wage cases. The only payoffs that can change are the ones for the teachers. These payoffs now depend on the number of students in the class.

(i) Both teachers solicit bribes. When both teachers solicit bribes, there are no differences compared to the fixed-wage case. Let n_{Gi} be the number of good students in the class of teacher T_i. The payoff for T_i is $\pi_i(n_{Gi}) = 40 + 4 \cdot 50 + 10 \cdot (4 - n_{Gi})$, which is the same as the payoff teachers obtain in the fixed-wage scenario with $F = 240$. Therefore, the expected utility for teachers who both solicit bribes is the same as in the fixed-wage with $F = 240$.

(ii) One teacher solicits bribes and the other teacher does not solicit bribes. Compared to the fixed-wage case, the payoffs for the two teachers change. Payoffs now depend on the number of bad students. Let n_B be the number of bad students. The teacher who solicits bribes (T_1) obtains a payoff $\pi_1(n_B) = 40 + n_B \cdot 50 + n_B \cdot 10$ and the teacher who does not solicit bribes (T_2) obtains a payoff $\pi_2(n_B) = 40 + (8 - n_B) \cdot 50$. The utility of T_1 conditional on n_B is

$$\begin{aligned}
U_{(1,0)}(n_G, n_B, \alpha, \beta) = \pi_1(n_B) + & \\
& - \beta \max(\pi_1(n_B) - \pi_2(n_B), 0) + \\
& - \beta \left[n_G \max(\pi_1(n_B) - \pi_G, 0) + (8 - n_G - n_B) \max(\pi_1(n_B) - \pi_M, 0) + n_B \max(\pi_1(n_B) - \pi_B, 0)\right] + \\
& - \alpha \max(\pi_2(n_B) - \pi_1(n_B), 0) + \\
& - \alpha \left[n_G \max(\pi_G - \pi_1(n_B), 0) + (8 - n_G - n_B) \max(\pi_M - \pi_1(n_B), 0) + n_B \max(\pi_B - \pi_1(n_B), 0)\right]
\end{aligned}$$

In this case, the unconditional expected utility is

$$EU_{(1,0)}(\alpha,\beta)=\sum_{n_G=0}^{8}\sum_{n_B=0}^{8-n_G}\frac{8!}{n_G!n_B!(8-n_G-n_B)!}\left(\frac{1}{6}\right)^{n_G+n_B}\left(\frac{4}{6}\right)^{8-n_G-n_B}U_{(1,0)}(n_G,n_B,\alpha,\beta)$$

that, after simplification, becomes

$$EU_{(1,0)}(\alpha,\beta)=120-\frac{2151595}{26244}\beta-\frac{22796875}{26244}\alpha$$

Similarly, one can define the conditional utility for the teacher who does not solicit bribes and compute the unconditional expected utility. After simplifications, the equation for the expected utility $EU_{(0,1)}(\alpha,\beta)$ takes the following form

$$EU_{(0,1)}(\alpha,\beta)=\frac{1120}{3}-\frac{183564625}{104976}\beta-\frac{206545}{104976}\alpha$$

(iii) Both teachers do not solicit bribes. This case differs slightly from the previous cases. Let n_G and n_B be the number of good and bad students. Note that, when both teachers do not solicit bribes, they share the good and intermediate students. Therefore, when a teacher has $\left\lceil\frac{8-n_B}{2}\right\rceil$ students in the class the other teacher has $\left\lfloor\frac{8-n_B}{2}\right\rfloor$ students in the class. The payoff of the former is $\pi_+(n_B)=40+\left\lceil\frac{8-n_B}{2}\right\rceil\cdot 50$ and the payoff of the latter is $\pi_-(n_B)=40+\left\lfloor\frac{8-n_B}{2}\right\rfloor\cdot 50$. Moreover, a teacher obtains $\pi_+(n_B)$ or $\pi_-(n_B)$ with equal probability. Therefore, the utility of a teacher conditional on n_B and n_G is either

$$\begin{aligned}U^+_{(0,0)}(n_G,n_B,\alpha,\beta)=\pi_+(n_B)+&\\
-\beta\max(\pi_+(n_B)-\pi_-(n_B),0)+&\\
-\beta\left[n_G\max(\pi_+(n_B)-\pi_G,0)+(8-n_G-n_B)\max(\pi_+(n_B)-\pi_M,0)+n_B\max(\pi_+(n_B)-\pi_B,0)\right]+&\\
-\alpha\max(\pi_-(n_B)-\pi_+(n_B),0)+&\\
-\alpha\left[n_G\max(\pi_G-\pi_+(n_B),0)+(8-n_G-n_B)\max(\pi_M-\pi_+(n_B),0)+n_B\max(\pi_B-\pi_+(n_B),0)\right]&\end{aligned}$$

with probability $\frac{1}{2}$ or

$$\begin{aligned}U^-_{(0,0)}(n_G,n_B,\alpha,\beta)=\pi_-(n_B)+&\\
-\beta\max(\pi_-(n_B)-\pi_+(n_B),0)+&\\
-\beta\left[n_G\max(\pi_-(n_B)-\pi_G,0)+(8-n_G-n_B)\max(\pi_-(n_B)-\pi_M,0)+n_B\max(\pi_-(n_B)-\pi_B,0)\right]+&\\
-\alpha\max(\pi_+(n_B)-\pi_-(n_B),0)+&\\
-\alpha\left[n_G\max(\pi_G-\pi_-(n_B),0)+(8-n_G-n_B)\max(\pi_M-\pi_-(n_B),0)+n_B\max(\pi_B-\pi_-(n_B),0)\right]&\end{aligned}$$

with probability $\frac{1}{2}$. Therefore, the unconditional expected utility is

$$\begin{aligned}EU_{(0,0)}(\alpha,\beta)=\sum_{n_G=0}^{8}\sum_{n_B=0}^{8-n_G}\frac{8!}{n_G!n_B!(8-n_G-n_B)!}\left(\frac{1}{6}\right)^{n_G+n_B}\left(\frac{4}{6}\right)^{8-n_G-n_B}\cdot&\\
\cdot\frac{1}{2}\left[U^+_{(0,0)}(n_G,n_B,\alpha,\beta)+U^-_{(0,0)}(n_G,n_B,\alpha,\beta)\right]&\end{aligned}$$

After simplifications, it becomes

$$EU_{(0,0)}(\alpha,\beta)=\frac{620}{3}-\frac{36176125}{139968}\beta-\frac{9115645}{139968}\alpha$$

Payoff matrix for the teachers under piece-rate.

After having obtained the utility for the teachers in all cases, we can write the payoff matrix for the piece-rate regime. When $F = 40$ and piece-rate of $s = 50$, we have the following payoff matrix for the teachers:

$$
\begin{array}{cc|c|c|}
 & & \multicolumn{2}{c}{T_2} \\
 & & B=1 & B=0 \\
\cline{3-4}
T_1 & B=1 & \frac{820}{3} - \frac{196754975}{209952}\beta - \frac{799775}{209952}\alpha & 120 - \frac{2151595}{26244}\beta - \frac{22796875}{26244}\alpha \\
\cline{3-4}
 & B=0 & \frac{1120}{3} - \frac{183564625}{104976}\beta - \frac{206545}{104976}\alpha & \frac{620}{3} - \frac{36176125}{139968}\beta - \frac{9115645}{139968}\alpha \\
\cline{3-4}
\end{array}
$$

Note that, conditional on T_2 soliciting bribes, T_1 prefers $B_1 = 1$ over $B_1 = 0$ when

$$\beta \geq \frac{466560}{3786095} + \frac{8593}{3786095}\alpha$$

and, conditional on T_2 not soliciting bribes, T_1 prefers $B_1 = 1$ over $B_1 = 0$ when

$$\beta \geq \frac{7278336}{14820571} + \frac{67480613}{14820571}\alpha$$

These inequalities identify regions with different equilibria of the game. These are reported in panel (c) of Figure 3 in the main paper.

C.2 Inequity averse students with the experimental parameters

Here we show that, for a constellation of parameters estimated in the literature (Goeree & Holt 2000, Blanco et al. 2011, Beranek et al. 2015), students' inequity aversion does not generate incentives to deviate from the selfish equilibrium strategy.

To check this, we verified that each type of student, i,e, bad, intermediate, and good, has no profitable deviation from the selfish strategy profile when assuming inequity aversion. The procedure is as follows: (i) we fix the type of one student and assume that the teachers and the other students follow the equilibrium strategy; (ii) we calculate the expected utility of this student for all his/her possible strategies;[12] (iii) we compare the expected utility of the equilibrium strategy and the utility of the alternative strategies.

Results of these calculations are reported in Table C.3. The table shows the results for the calculations for a range of estimated inequity aversion parameters, and for the following three cases: (i) Fixed Wage when teachers do not solicit bribes; (ii) Piece rate when both teachers do not solicit bribes; and (iii) Piece rate when both teachers solicit bribes. Comparing the expected utility of the equilibrium strategy to the expected utility of the best alternative strategy, we show that, in all these cases, students do not have an incentives to deviate.[13]

[12]For each strategy, we obtained the expected utility by calculating the utility over all the 128 possible combination of other students' types, multiplying each of these for the probability to observe that distribution, and taking the sum.

[13]Note that the equilibrium is not a strict equilibrium in case (iii)

Table C.3: Students' incentive to deviate from the equilibrium with Inequity Averse teachers. Estimated expected utility for the three cases.

Paper	BEN (2011) $\alpha = 0.910; \beta = 0.380$		GH prop (2000) $\alpha = 0.860; \beta = 0.660$		GH resp (2000) $\alpha = 0.860; \beta = 0.120$		BCG Nott. (2015) $\alpha = 0.754; \beta = 0.484$		BCG Izmir (2015) $\alpha = 1.227; \beta = 0.589$		BCG M-Turk (2015) $\alpha = 1.218; \beta = 0.410$	
(i) Fixed Wage when both teachers do not solicit bribes												
Student type	EU eq	EU best alt.	EU eq	EU best alt.	EU eq	EU best alt.	EU eq	EU best alt.	EU eq	EU best alt.	EU eq	EU best alt.
Bad	-542.475	-772.225	-506.350	-724.850	-506.350	-724.850	-429.765	-624.415	-771.507	-1072.582	-765.005	-1064.055
Intermediate	-4.525	-542.475	-18.683	-506.350	25.417	-506.350	14.152	-429.765	-76.804	-771.507	-60.618	-765.005
Good	87.050	-542.475	-25.650	-506.350	191.700	-506.350	45.190	-429.765	2.928	-771.507	74.975	-765.005
(ii) Piece Rate when both teachers do not solicit bribes												
Student type	EU eq	EU best alt.	EU eq	EU best alt.	EU eq	EU best alt.	EU eq	EU best alt.	EU eq	EU best alt.	EU eq	EU best alt.
Bad	-444.533	-719.142	-413.940	-674.684	-413.671	-674.684	-348.698	-580.432	-639.486	-1001.008	-633.865	-993.005
Intermediate	41.837	-444.533	23.564	-413.940	70.477	-413.671	51.685	-348.698	-14.691	-639.486	1.950	-633.865
Good	64.883	-444.533	-64.150	-413.940	184.700	-413.671	16.957	-348.698	-31.431	-639.486	51.058	-633.865
(c) Piece Rate when both teachers solicit bribes												
Student type	EU eq	EU best alt.	EU eq	EU best alt.	EU eq	EU best alt.	EU eq	EU best alt.	EU eq	EU best alt.	EU eq	EU best alt.
Bad	-119.191	-119.191	-112.448	-112.448	-97.759	-97.759	-77.217	-77.217	-215.920	-215.920	-208.466	-208.466
Intermediate	-119.191	-119.191	-112.448	-112.448	-97.759	-97.759	-77.217	-77.217	-215.920	-215.920	-208.466	-208.466
Good	-32.885	-32.885	-49.170	-49.170	0.805	0.805	-12.472	-12.472	-113.264	-113.264	-94.965	-94.965

D. Additional analysis [ONLINE APPENDIX]

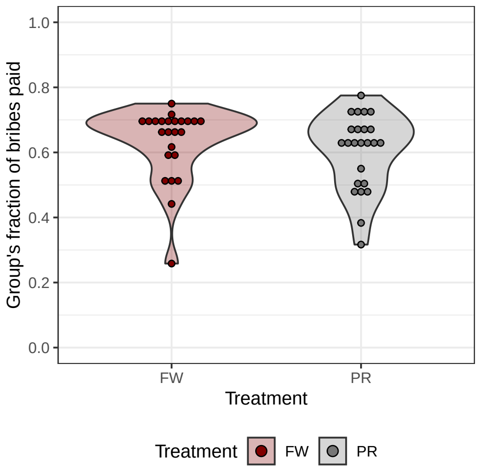
D.1 Overall bribing

Figure D.1 reports information regarding fraction of bribes paid. Note that bribes are not paid either because the students did not have the opportunity to pay or because the student had the opportunity and decided not to pay the bribe. Results are similar to the ones obtained looking at the fraction of teachers soliciting bribes, which are reported in the paper. The fraction of bribes paid in part 1 does not differ across the two treatments. A Wilcoxon rank sum test does not reject the null hypothesis that the probability of a random observation from one treatment exceeding a random observation from the other is one half. In part 2, the fraction of bribes paid is significantly lower in PR compared to FW (Wilcoxon rank sum test $p < 0.001$). As for the comparison of part 1 and part 2, we find a significant difference in overall bribing in PR (Wilcoxon signed rank test $p < 0.001$) but not in FW (Wilcoxon signed rank test $p = 0.170$).

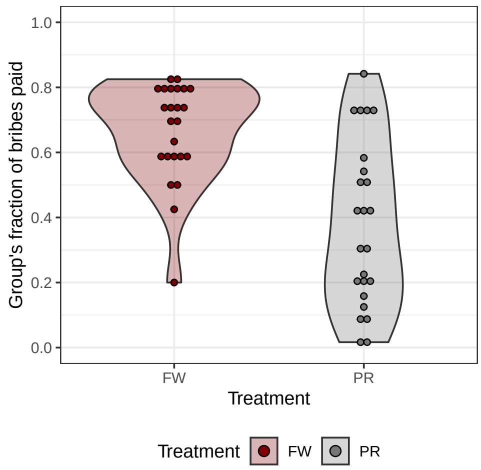
Panel (c) of Figure D.1 reports the theoretical fraction of bribes that should have been paid given the choice of the teachers and the ability of the students (crosses), along with the observed fraction of bribes paid (dots). With fixed-wage the actual frequency of bribes paid is slightly lower than the theoretically predicted fraction.

Figure D.1: Group's fraction of bribes paid in part 1 (a) and in part 2 (b) and fraction of students paying bribes over periods (c)

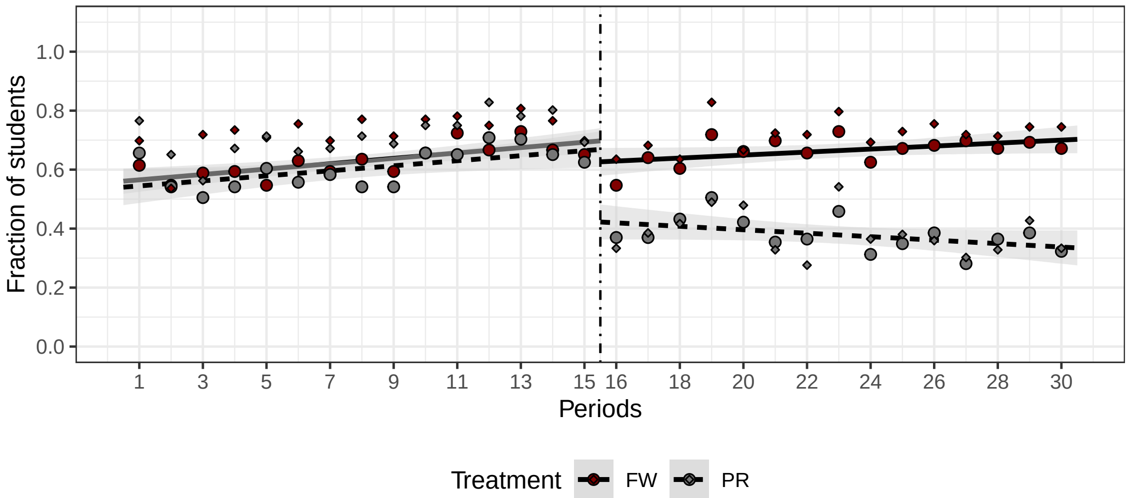
(a) Part 1



(b) Part 2



(c) Fraction of bribes paid over periods (diamonds are the predicted fractions when students follow the equilibrium strategy after observing the decisions of the teachers)



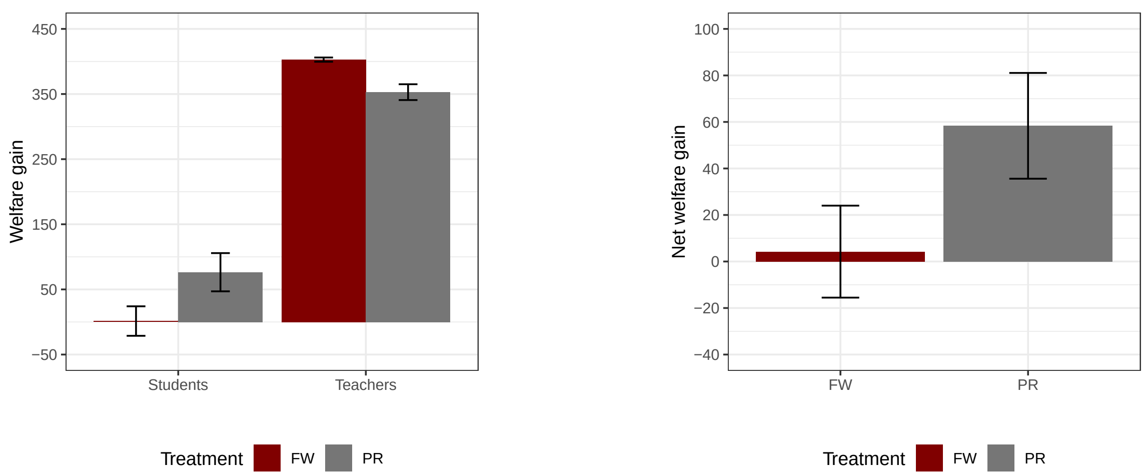
D.2 Social welfare

We analyze who benefits the most from the two interventions by comparing the increase in total payoffs for teachers and for students when moving from part 1 to part 2. As is apparent from Figure D.2 panel (a), which depicts the increase in the group total welfare for teachers and students by treatment, the introduction of a high fixed-wage particularly benefits teachers. They receive about 400 extra points per period (Wilcoxon signed rank test $p < 0.001$), while the students do not benefit at all (Wilcoxon signed rank test $p = 0.900$). The introduction of the piece-rate bonus, on the other hand, benefits both teachers and students who enjoy an extra payoff of about 350 and 80 points per period, respectively (Wilcoxon signed rank test $p < 0.001$ in both cases).

Figure D.2: Welfare difference between part 2 and part 1 for students and teachers by treatment (panel (a)) and total efficiency gain between part 1 and part 2, i.e., increase in group total payoff net of the public expenditure, by treatment (panel (b)). (Mean and 95% C.I.)

(a) Welfare gain by role

(b) Net welfare gain



Although students overall prefer the introduction of the piece-rate bonus over fixed-wage increase (Wilcoxon signed rank test $p < 0.001$ in both cases), differences in the preference across student types exists. That is, while the piece-rate scheme benefits students with low and medium effort costs, the minority of students with high effort costs prefer to obtain their diploma by paying a bribe (and not exerting effort). Teachers, on the other hand, would prefer the introduction of a high fixed-wage over a piece-rate bonus (Wilcoxon rank sum test $p < 0.001$). The reason lies in the disincentive to solicit bribes in the piece-rate regime which, in turn, produces an incentive to choose the outside option for the students with high effort costs. Hence, the total wage paid to the teachers becomes smaller. This, however, implies that the piece-rate bonus policy is on average less expensive.

Finally, to estimate the overall welfare benefits of the two interventions we account for the different public expenditure levels. Figure D.2 panel (b) illustrates that the fixed-wage increase fails to improve net welfare when moving from part 1 to part 2 (Wilcoxon signed rank test $p = 0.705$). The introduction of the piece-rate scheme, on the other hand, produces a significant welfare gain (Wilcoxon signed rank test $p < 0.001$). Taken together, the new results suggest that introducing a piece-rate scheme in corrupt education systems can help to reduce the occurrence of bribe transactions and improve overall social welfare, in particular for students (with low and medium effort costs).

D.3 Dynamic play of teachers

Here we look at the dynamics of teachers' behavior in the second part of the experiment. This analysis permits gaining deeper insights into how the shift from a corrupt system to one with less corruption occurs. In particular, we study how teachers react to their own choices, as well as, to the other teachers' choices. Table D.1 reports the fraction of times a teacher solicited bribes conditional on his/her choice, and the other teacher's choice in the previous period. In the *FW* treatment, teachers tend to solicit bribes independent of the previous period's choices. This is in line with selfish preferences and not with inequity aversion. In the *PR* treatment, instead, teachers tend to solicit bribes if the other teacher did so in the previous period (more strongly if both did so) and tend to abstain from soliciting bribes if the other teacher also abstained. This is not in line with the predictions based on selfish preferences. Instead, in combination with the pattern observed in treatment *FW*, it suggests that the dynamics are in line with a tit-for-tat type of behavior, which is compatible with incomplete conditional cooperation (Fischbacher & Gächter 2010).

Table D.1: Teachers' fraction of "solicit" choices conditional on previous period decisions (part 2 data); + means that behavior is in agreement with a model; - that it is in disagreement

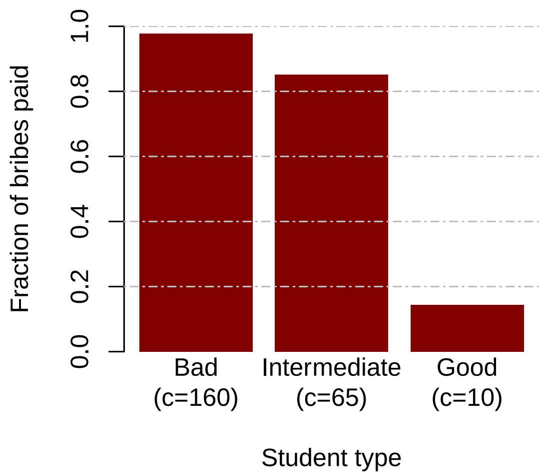
	Both do not solicit	Solicit and the other does not	Do not solicit and the other does	Both solicit
Piece-Rate	26.3%	33.9%	54.1%	83.7%
Inequity Aversion	+	+	+/−	+
Selfish Preferences	+	+	+/−	−
Tit-For-Tat	+	+	+/−	+
Fixed-Wage	75.0%	71.2%	88.1%	93.4%
Inequity Aversion	−	−	−	−
Selfish Preferences	+	+	+	+
Tit-For-Tat	+	+	+	+

D.4 Students individual behavior

D.4.1 Choice to pay the bribe or not

Figure D.3 shows the fraction of bribes paid by effort cost. These frequencies are close to the predicted frequencies of 1, 1, and 0 for the bad, the intermediate, and the good students, respectively.

Figure D.3: Fraction of bribes paid by effort level



Deviations from the equilibrium predictions are very rare when they are extremely costly, i.e., not paying the bribe when the effort cost is high, and they are more common when the cost of deviating is smaller, i.e., in case of medium and low effort cost.

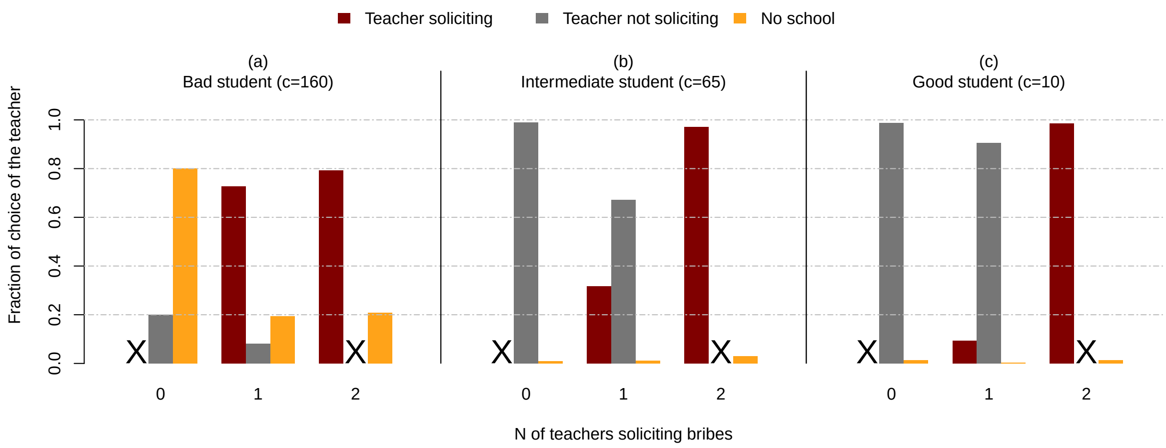
D.4.2 Choice to go to school

Figure D.4 reports the students' choice of the teacher by effort cost and by number of teachers who solicit bribes. When the number of teachers soliciting bribes is 0, the choice is between the "Teacher not soliciting" and "No school". When the number of teachers soliciting bribes is 2, the choice is between the "Teacher soliciting" and "No school". When the number of teachers soliciting bribes is 1, the choice is among all three available options.

Recall that the equilibrium predictions are as follows:

- When the number of teacher soliciting bribes is 0, bad students should choose "No school" and the other students should choose "Teacher not soliciting".
- When the number of teacher soliciting bribes is 1, bad students should choose "Teacher soliciting" and the other students should choose "Teacher not soliciting".
- When the number of teacher soliciting bribes is 2, all students' types should choose "Teacher soliciting".

Figure D.4: Students' choice of the teacher by student's type cost and options available



As for good students, panel (c) in Figure D.4 illustrates that the observed fractions of choices are largely in line with the equilibrium predictions. The highest deviation rate occurs when one teacher solicits bribes and the other teacher does not solicit bribes. In this case, 9.5% of the choices deviate from the ones predicted by the equilibrium.

For the intermediate students, choices are in line with predictions when both teachers make the same choice (0 and 2). When one teacher solicits and the other does not solicit bribes, about one third of the choices are for "Teacher soliciting" (31.6%). This is not compatible with the equilibrium predictions, but it can be profitable if other students do not pay the bribe when they should. Recall that "Teacher not soliciting" guarantees a payoff of 185. "Teacher soliciting", instead, can provide a higher amount only when paying the bribe and some other students do not pay the bribe. Empirically, choosing a "Teacher soliciting" and behaving optimally (pay when more than 2 students are in the class and not paying when alone in the class) gives a payoff $\geq$ 185 only 46 out of 425 times (10.82%).

As for the bad students, when "Teacher soliciting" is available, 20% of the choices are for "No school". This may be due to a preference for not paying the bribe. Indeed, these students are willing to give away about 25 points to avoid paying the bribe to the teacher (In case of a fully corrupt class of 4 students, choosing "Teacher soliciting" and paying gives 140 points and "No school"gives 115). Moreover, about 20% of the choices are for "Teacher not soliciting" when the "No school" option is available (payoff of 90 instead of 115).

D.5 Moral vs equilibrium play

Previous results show that students' behavior is close to the equilibrium predictions most of the time. Here, we look at whether students decisions can be explained by moral concerns. To do so, we compare two possible styles of play for the students: equilibrium play and moral play. Moral play is defined by choosing the best strategy from the set that excludes paying a bribe. Table D.2 summarizes the predicted choices for all possible situations for the two styles of play.

Table D.2: Summary of the choices for the different styles of play (N.S.= no school; B.F.S. = bribe free school (teacher not soliciting); B. S. = bribe school (teacher soliciting))

N. teachers soliciting bribes	Effort Cost	Equilibrium	Moral	Mistake
0	High	N.S.	N.S.	B.F.S.
1	High	B.S. + pay	N.S.	B.F.S. B.S. + not pay
2	High	B.S. + pay	N.S.	B.S. + not pay
0	Medium	B.F.S.	B.F.S.	N.S.
1	Medium	B.F.S.	B.F.S.	N.S. B.S. + pay B.S. + not pay
2	Medium	B.S. + pay	B.S. + not pay N.S.	—
0	Low	B.F.S.	B.F.S.	N.S.
1	Low	B.F.S.	B.F.S.	N.S. B.S. + pay B.S. + not pay
2	Low	B.S. + not pay	B.S. + not pay N.S.	B.S. + pay

Figure D.5 panel (a) shows the fraction of periods where choices are consistent with equilibrium play and moral play for each individual. Note that the sum of the fractions can be greater than 1. This is due to the fact that some choice patterns are compatible with both equilibrium play and moral play. In general, equilibrium play seems to better predict behavior—most of the individuals are in the bottom right corner—. The subjects with a fraction of equilibrium choices strictly higher than the fraction of moral choices are 348 out of 384.

Figure D.5: Students' individual fraction of choices in line with equilibrium play (x) and with moral play (y). Points represents participants. Panel (a) uses all choices and panel (b) only the choices where the predictions differ.

(a) All choices

(b) Moral $\neq$ equilibrium

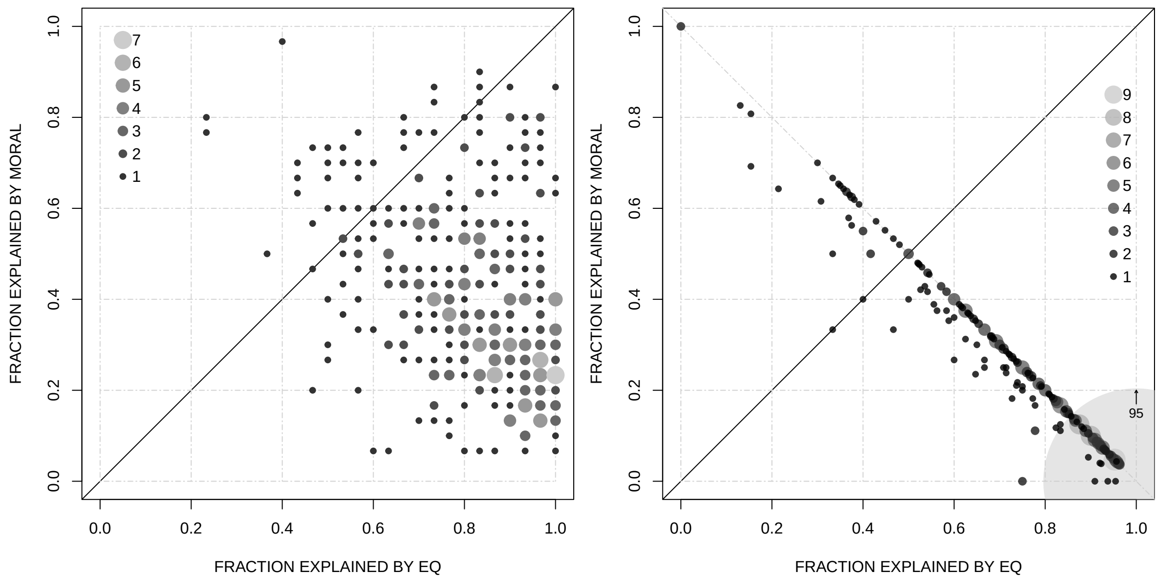


Figure D.5b panel (b) takes into account the fact that the predictions of the two styles may overlap. Therefore, the Figure considers only the data of those situations in which equilibrium play and moral play provide different predictions—i.e., when effort cost is low and there is at least one teacher soliciting bribes and when effort cost is medium and both teachers solicit bribes. In contrast to the panel (a), the sum of the fraction of choices explained by the two styles cannot be greater than one. Figure D.5b panel (b) illustrates that equilibrium play seems to better predict behavior (most of the individuals are in the bottom right corner). In this case, the subjects with a fraction of equilibrium choices strictly higher than the fraction of moral choices are 349 out of 384.

As a second step, we look at differences in the fraction of the two styles of play across treatments. Figure D.6 and Figure D.7 replicate the analyses in Figure D.5b for the PR and for the FW treatment separately. Comparing the number of subjects with a fraction of equilibrium choices strictly higher than the fraction of moral choices in the two treatments does not reveal significant differences in the style of play under piece-rate versus under fixed-wage (170 subjects out of 192 in PR and 179 out of 192 in FW; $\chi^2(1) = 2.546, p = 0.111$). Analyzing the number of subjects that consistently play according equilibrium predictions leads to the conclusion (45 subjects out of 192 in PR and 50 out of 192 in FW; $\chi^2(1) = 0.350, p = 0.554$).

Figure D.6: Students' individual fraction of choices in line with equilibrium play (x) and with moral play (y) in the PR treatment. Points represents participants. Panel (a) uses all choices and panel (b) only the choices where the predictions differ.

(a) All choices

(b) Moral $\neq$ equilibrium

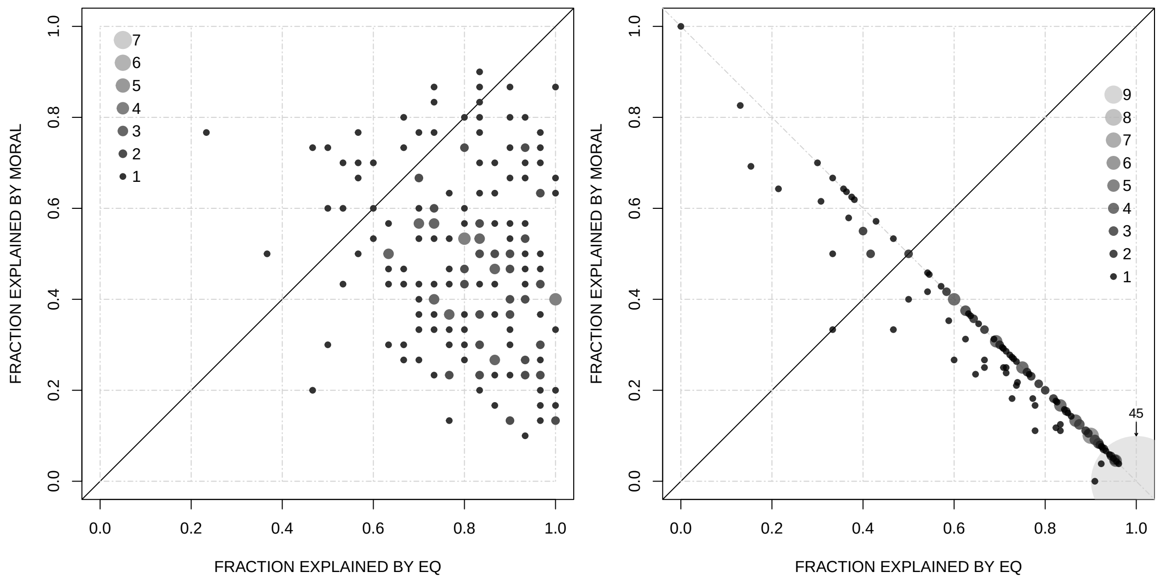


Figure D.7: Students' individual fraction of choices in line with equilibrium play (x) and with moral play (y) in the FW treatment. Points represents participants. Panel (a) uses all choices and panel (b) only the choices where the predictions differ.

(a) All choices

(b) Moral $\neq$ equilibrium