A market for integrity
An experiment on corruption in the education sector

Ivan Soraperra*[1], Nils Köbis†[1], Charles Efferson‡[2], Shaul Shalvi§[1], Sonja Vogt¶[3,4], and Theo Offerman‖[1]

[1]CREED - University of Amsterdam
[2]HEC Lausanne, University of Lausanne
[3]CESS - University of Oxford
[4]University of Bern

June 18, 2019

Abstract

Corruption in the education sector is pervasive in many (developing) countries. We examine two interventions to fight corruption in education. The first is an increase of the fixed-wage of teachers. The second is the introduction of a piece-rate scheme that rewards teachers according to the number of students that they attract. We model these mechanisms and conduct a lab experiment in Colombia, a country riddled with corruption. After creating a culture of corruption, we introduce either intervention. The increase of the fixed-wage does not diminish bribery. The piece-rate scheme substantially reduces but does not eliminate bribery.

Keywords: Education, corruption, theory, experiment
JEL Classification: D73, D82, C90

*Shared first authorship. Email: i.soraperra@uva.nl
†Shared first authorship. Email: n.c.kobis@uva.nl
‡Email: charles.efferson@unil.ch
§Email: s.shalvi@uva.nl
¶Email: sonja.vogt@soz.unibe.ch
‖Email: t.j.s.offerman@uva.nl

1 Introduction

Competition increases unethical conduct such as corruption—a prominently voiced (Shleifer, 2004) and widely shared view in economics and beyond (Falk and Szech, 2013). Especially forms of corruption that entail public officials pocketing government funds—i.e. corruption with theft (Shleifer and Vishny, 1993)—flourish when rule of law is weak and competition is fierce. Competition fosters corruption among private firms that acquire public permits by undercutting official prices through bribery. Also within public administration, competition for lucrative public posts that allow rent extraction places a costly premium on integrity. Although the notion widely applies, we present the case of systemic corruption in the education sector where competition can help to reduce corruption. It provides a salient example where the conceived wisdom of competition increasing corruption is overturned.

Corruption in the education sector presents an immense challenge for societies around the world, being particularly pronounced in many developing countries in Africa, Eastern Europe, Asia and South-America. Widespread practices such as bribery between students and teaching staff undermine equal opportunities for knowledge acquisition (Rothstein, 2011)—which has been shown to have far reaching consequences on non-economic outcomes such as health in adulthood (Conti et al., 2010)—, reinforce socio-economic inequalities (Hallak and Poisson, 2007), and "corrupt" young people's concepts of fairness (Heyneman et al., 2008), leading to emigration of talent (Cooray and Schneider, 2016). According to the Global Corruption Barometer (TI, 2013), many survey respondents who came into contact with education services reported to have paid bribes (e.g. 48% in India, 75% in Liberia and 62% in Sierra Leone). Although substantial amounts of money are invested into improving education in developing countries—the EU plans to spend 4.7 billion Euro in the period 2014-2020—these (financial) efforts are impeded by corruption (Baldacci et al., 2008).

In this paper, we introduce a novel theoretical framework that models one of the most pervasive forms of corruption in the education sector, namely bribery transactions between students and teachers. We study the possibility of corruption in the absence of reliable punishment institutions—reflecting the common situation of teachers' impunity in many developing countries (Cadot, 1987). When teachers solicit bribes, students face a social dilemma (Rothstein, 2011). 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.

In this setup, we test the effectiveness of two interventions for education systems riddled with high levels of corruption. Theoretically and in an experiment, we compare the commonly proposed policy of a fixed-wage increase for teachers (Rose-Ackerman and Palifka, 2016; Tanzi, 1998; Fisman and Golden, 2017) to a new alternative of a piece-rate scheme to fight corruption. The idea behind the former intervention is that teachers have to accept bribes to increase their meagre salary that is insufficient to support their families. Grounded in Akerlof and Yellen's (1990) fair-wage hypothesis, the thinking is that teachers can only afford behaving morally when their base wage is set at an adequate level. The latter intervention consists of a piece-rate scheme, that pays teachers according to the number of students that they attract (Friedman, 1955, 1962; Hoxby, 1994, 2000). The logic behind this novel approach to fight corruption: schools will be forced to curb bribery levels when they have to compete for students. If many students acquire a diploma by bribery instead of hard work, the value of a diploma will be eroded.[1] Many students in turn avoid schools whose diplomas have lost most of their reputational value because of bribery. By voting with their feet, students may force teachers to stop soliciting bribes, hence representing an endogenous punishment mechanism (Sutter et al., 2010).

In Section 2, we will present the available empirical evidence for both interventions and show that it is inconclusive. Whereas empirical evidence on the effectiveness of the fixed-wage increase is mixed, empirical evidence on the success of piece-rate schemes to fight corruption is lacking altogether.

So far, no research has assessed the effectiveness of the two schemes when introduced under identical circumstances. To fill this gap, we conducted a large pre-registered lab experiment in Colombia, a country that experiences high levels of corruption.[2] A major advantage of lab experiments is that they allow to directly measure otherwise unobservable behavior, such as corruption, in a controlled environment (Falk and Heckman, 2009).

The experiment implements our corruption game and tests some of its predictions. The predictions depend on whether players are selfish or have other-regarding preferences. With selfish preferences, a fixed-wage regime will push the welfare of the teachers to a higher level without affecting the marginal incentives. Thus, teachers will take the higher wage and continue to solicit bribes. The piece-rate regime, instead, disciplines teachers if the subsidy is sufficiently high. To attract those students who prefer to acquire a valuable diploma by working hard, teachers have an incentive to stop soliciting bribes.

We further analyze the game with social preferences assuming inequity aversion (Fehr and Schmidt, 1999), which provides the simplest theoretical account to support the success of a fixed-wage increase. When teachers are sufficiently inequity averse, a flat salary increase will eradicate corruptible behavior of teachers. Thus, inequity aversion can support Akerlof and Yellen (1990)'s fair-wage hypothesis. This is one of the possible motives that can sustain the hypothesis, for an account of different fairness motives see, e.g., Cappelen and Tungodden (2019). Whether the adoption of a piece-rate regime achieves the intended goal depends on the equilibrium on which teachers focus. Both mutual non-corruptibility and mutual corruptibility are supported in equilibrium.

In our experiment, we first create a culture of corruption by paying low fixed-wages to teachers. After high levels of bribery stabilize, either a fixed-wage or a piece-rate salary scheme is introduced. While bribery levels remain high after the introduction of a substantial fixed-wage increase, the piece-rate salary scheme substantially reduces corruption—creating a market for integrity.[3]

[1] A salient example is the PhD degree obtained by Grace Mugabe, the former first lady of Zimbabwe. She was awarded the doctorate in sociology by the University of Zimbabwe (UZ) in August 2014, just three months after she started. This case attracted a lot of scorn worldwide, e.g. https://www.nytimes.com/2017/01/07/world/africa/in-zimbabwe-a-first-lady-exerts-her-power.html

[2] For pre-registration, see Open Science Framework.

[3] 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 to get caught and punished for bribing is practically zero; 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 and Schargrodsky, 2003; Lambsdorff and Fink, 2006). The proposed piece-rate

We regard the evidence in this paper as a proof of concept. A natural question that we will further address in Section 6 is whether and how the results of our experiment can be extrapolated to the field. Here, we note that Armantier and Boly (2013) provide supportive evidence that lab and field data on corruptive practices in the education sector are congruent. They compare how graders of exams respond to bribing requests in the lab in Canada, and in the lab and in the field in Burkina Faso. They report that 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 the more promising avenue to reduce corruption.

The remainder of the paper is structured as follows. Section 2 reviews the literature on whether corruption within the education sector can be reduced by higher wages. Section 3 introduces our theoretical model. Section 4 discusses the experimental design and specifies the predictions. Section 5 presents the experimental results. Section 6 discusses theoretically how sensitive the results are to changes in the environment. Section 7 concludes.

2 Literature on (fighting) corruption in education

Large global surveys on corruption and education commissioned by Transparency International indicate that overall 41% of the respondents perceive their respective education sector to be corrupt or highly corrupt—exceeding 70% in several countries such as Russia, the Democratic Republic of Congo or Kyrgyzstan ($N = 114,000$ in 107 countries, TI, 2013).

Among the many forms of corruption in the education sector, bribery between teaching staff and students marks one of the most prevalent and problematic forms of corruption (UNDP, 2011; TI, 2005). Survey evidence suggest that 60% to 80% of students in Eastern European states pay bribes to enter or advance in the education system (Denisova-Schmidt and Prytula, 2017).[4]

2.1 Empirical evidence on fair-wage hypothesis

To reduce corruption in education, the literature is unanimous in attributing a high importance to adequate salaries for teaching staff (Dolan et al., 2012; UNDP, 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 and Stiglitz, 1984). According to Akerlof and 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.

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 and Golden, 2017; Rothstein, 2011).

[4]Ferraz et al. (2012) provide further evidence by using data based on audit reports. Different corruption levels across Brazil's municipalities suggest immense efficiency costs on learning outcomes: students in municipalities riddled with corruption score 0.35 standard deviations lower on test scores compared to students in relatively corruption-free municipalities. Higher corruption was also related to higher drop-out and failure rates (see for further discussion Olken, 2009)

In order to gauge whether higher salaries indeed lead to lower levels of corruption, we conducted a pre-registered systematic literature review (see more details on Open Science Framework). Overall, 3764 articles were identified. All of them were inspected to see whether they met pre-specified inclusion criteria. In total, 57 studies are included in the qualitative synthesis, and they paint a rather mixed picture: 17 show an inverse relationship (i.e., an increase in salary leads to a decrease in corruption), 33 studies show no or a mixed relationship, while 7 articles reveal a positive association (i.e., an increase in salary is associated with an increase in corruption). This systematic literature review is presented in Appendix A.[5]

The available empirical evidence suggests that raising public salaries without accompanying monitoring mechanisms jeopardizes the effectiveness of salary increases—at times even leading such policies to backfire, potentially because these policies create exaggerated expectations of rent-extraction opportunities (Cappelen et al., 2018). Historically some cases exist that provide support for the assumed link between higher public wages and a reduction in corruption. Two prominent cases are the transformations of Hong Kong and Singapore which saw public salary increases schemes as well as a reformation of the policing institutions that subsequently led to unprecedented drop in corruption levels (Svensson, 2005). In the same spirit, Borcan et al. (2014) find that a wage *cut* leads to *more* corruption. 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 in which the teachers' salaries remained constant.

In many 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. (2015). They report results from a large-scale randomized experiment in which the wage increase of teachers was accelerated in the treated schools. They find 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. Relatedly, Mishra et al. (2008) study the effect of the Indian tariff reform in the 1990s. Their results show that bribery persisted among Indian customs inspectors, even after a fixed monthly salary increase of up to 80-100 percent.

There is also evidence that an increase in the wages of public officials may backfire. Foltz and Opoku-Agyemang (2015) investigate the impact of a Ghanaian wage reform that saw the police force's salary being doubled in 2010. They find that the salary increase led to higher levels of effort to collect bribes by the police officers, who operated more checkpoints. The salary increase led to higher bribes being requested, and an increase in the total *value* of bribes paid, even though the total *number* of bribes paid decreased.

[5] This picture is confirmed in some correlational studies. Van Rijckeghem and Weder (2001) conducted an international comparative study to investigate the relationship between public wages and perceived levels of corruption for 31 developing countries and low income OECD countries. They find a weak but robust negative relationship between public sector wages and perceived levels of corruption, even when controlling for national level GDP, indicators of "rule of law" and "quality of government". Other cross-national studies on the subject by Rauch and Evans (2000) and Treisman (2000) find no robust correlation between a salary increases and corruption. More recently, An and Kweon (2017) use cross-country data of a time span of 10 years and find a weak negative relationship between public sector wages and perceived corruption. Naturally, these studies shed no light on the causality between public wages and corruption.

The rather 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 of corrupt behavior (Jacquemet, 2005; Barr et al., 2009)—in particular when wage increases are combined with effective punishment (Van Veldhuizen, 2013; Azfar and Nelson, 2007). Of special importance for our study are the experimental papers by Armantier and Boly (2011, 2013), who combine lab and field experiments to study bribery in the education system in Burkina Faso. In Armantier and 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 less mistakes. In a later version with the same experimental set-up, Armantier and 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 effect was statistically indistinguishable across the three settings.

Taken together, the results from lab and field experiments on salaries and corruption suggest that higher salaries can reduce corrupt behavior. However, this negative link specifically occurs when realistic threats of punishment exist.

2.2 Empirical evidence on piece-rate hypothesis

Empirical evidence on the success of piece-rate schemes to fight corruption in developing countries is lacking altogether. Related evidence stems from the introduction of voucher systems in education. In the case of the nation-wide voucher system introduced in Chile the intervention proved rather unsuccessful in improving educational outcomes (Hsieh and Urquiola, 2006)—one of the main reasons being the insufficient size of the subsidies (Verger et al., 2016; Chumacero et al., 2011). Additional evidence stems from the US where lottery programs award vouchers to disadvantaged students to cover tuition at eligible private schools. The success of such programs can be evaluated by comparing outcomes for lottery winners and losers. Most papers report either small positive or zero effects of lottery programs on student achievement (see the literature review in Abdulkadiroğlu et al., 2018). Abdulkadiroğlu et al. (2018) provide a striking example where a lottery program yields a negative effect. In the Louisiana Scholarship Program, which entails public funds to attend private schools, the results reveal that the vouchers reduced academic achievements.

Our conjecture is that an important reason behind the lack of success of a voucher system in Chile and the mixed support for the effectiveness of lottery programs for charter schools in the US is the limited size of the subsidies. For instance, it may be extremely costly to introduce subsidies that would make high-quality private schools interested in competing (Abdulkadiroğlu et al., 2018). The introduction of a piece-rate regime may be more effective in developing countries where the base level of wages is relatively low. This feature is supported by our model, outlining that a subsidy encourages competition for students and good behavior, yet only if it surpasses a threshold. Although controversy around voucher systems in the education sector exists, we argue that enabling students and their parents to reward teachers' integrity behavior can provide a useful new approach for anti-corruption in contexts where introducing punishment is unfeasible (see for reviews on the failed attempts to fight corruption by installing punishment institutions, Mungiu-Pippidi, 2017; Fisman and Golden, 2017; Heywood, 2018).[6]

3 Theoretical model

In this Section, we present the core model of bribery in the education sector, assuming selfish preferences. In Section 4.2.2, we discuss how other-regarding preferences can support the fair-wage hypothesis. In Section 6.1 we investigate how the predictions vary when some of the assumptions are modified. To model corruption in education we analyze the interaction between 2 teachers (or schools)—call them T_1 and T_2—and N students with heterogeneous abilities. Teachers decide whether or not to solicit bribes from students. Students observe the choices of the teachers and choose between three options, (i) they can choose to study with T_1; (ii) they can choose to study with T_2 or (iii) they can choose not to go to school (NS). When they join a teacher who solicits bribes, students choose whether or not to pay the bribe. The timing of the events is as follows:

- The N students are privately informed of their ability, i.e., of their effort cost c. The type of each student is independently drawn from a distribution with cdf $\mathcal{F}(c)$ and support $[0, \bar{c}]$.
- Teachers decide simultaneously whether or not to solicit bribes; they choose between "Bribe" ($B = 1$) and "No-Bribe" ($B = 0$).
- Students are informed about the teachers' decisions and simultaneously make their choice about the school. The options they can choose from depend on the teachers' decisions. If the teachers made different choices, i.e. one teacher is soliciting bribes while the other is not, the students can choose one of the following options: (i) to study with the teacher soliciting bribes, (ii) to study with the teacher not soliciting bribes, or (iii) not to study at all and go for the outside option (NS). If the teachers instead made the same choice, i.e. they are both soliciting or both not soliciting bribes, the students can choose one of the following options: (i) to study with one of the teachers or (ii) not to study at all and go for the outside option (NS). In this case, the students choosing to study are randomly

[6]Other ways to fight corruption in the public sector include rewarding of integrity (OECD, 2018). By encouraging pioneers to deviate from corruption these systems aims to reward "norm entrepreneur" and create "islands of integrity" which in turn can potentially trigger a cascade of corruption norms change (Bicchieri, 2016). Banerjee et al. (2008) show that the success of incentive schemes to reduce absenteeism in the public health service in India hinges on reliable monitoring. Initially an effective system was in place that led to less corruption among nurses in a response to incentives. Gradually, the positive effect vanished when monitoring was undermined. Duflo et al. (2012) tested an incentive scheme together with tamper-prove cameras in 57 randomly selected schools in India. Teachers received rewards for each day they were recorded to be present which affected absenteeism of teachers and students' test scores positively.

split between the two teachers.[7]

- Students are informed about the number of students choosing the two schools. The students who are in a school where bribes are solicited, simultaneously decide whether to pay the bribe ($P = 1$) or reject the request and not pay the bribe ($P = 0$). Students who are in a school where bribes are not solicited do not pay a bribe ($P = 0$).

3.1 Payoffs for the students

By studying hard, i.e. by paying their effort cost, students can obtain a diploma that is worth at most a standardized payoff of 1. The value of the diploma on the market, however, is lower when the reputation of the school is damaged. The higher the fraction of bribing students in the school, the lower is the reputation and the lower is the value of the diploma. We assume that the cost of reputation is proportional to the fraction of students bribing in the school. Therefore, if k is the number of students bribing in a school with n students, the value of the diploma is reduced by $\frac{k}{n}a$. This means that, independently of the school size, the value of the diploma is reduced by a when all the students pay the bribe and it is not reduced when all the students do not pay the bribe.

A student who studies hard to get the diploma incurs an effort cost c that depends on the student's ability. In case the teacher is accepting bribes, the student can pay a bribe $b > 0$ and obtain the diploma without paying the effort cost c.

The payoff for the student $\pi(P, n_b, n, c)$ depends on: the student's choice to pay the bribe P, on the number of other students in the school n, on the number of other students bribing the teacher in the same school n_b, and on the student's ability c. A student's payoff who does not pay the bribe to the teacher ($P = 0$) is:

$$\pi(0, n_b, n, c) = 1 - c - \frac{n_b}{n}a$$

A student who decides to pay the bribe ($P = 1$) earns:

$$\pi(1, n_b, n, c) = 1 - b - \frac{1}{n}a - \frac{n_b}{n}a$$

Students have the outside option of not going to school, i.e., NS. In this case they obtain a fixed payoff of π_{out}. When the student is in a class where bribes are not solicited no-one can pay the bribe and the payoff is $\pi(0, 0, n, c) = 1 - c$.

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. 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 the education sector in many developing countries. First, the market is aware when the diploma of a school suffers from reputation costs. Employers and universities can judge the quality of the students coming from different

[7] If n is the number of students that decide to go to school, one teacher randomly gets $\lfloor \frac{n}{2} \rfloor$ of the students and the other teacher randomly gets $\lceil \frac{n}{2} \rceil$ of the students.

schools and can discriminate among them (Heyneman et al., 2008). This implies that a diploma from a corrupt institution harms even the hard-working students. Second, students and their families are well aware of the quality of the teaching offered at an institution (Chumacero et al., 2011). In particular, they know which institutions accept and which do not accept bribes.

3.2 Payoffs for the teachers

The payoffs of the teachers depend on the institutional setting. In *fixed-wage* the teacher obtains a fixed payment of F and in *piece-rate* the teacher obtains a piece-rate payoff s for each student that chooses the teacher's school. In addition, independently of the setting, the teacher collects the bribe whenever a student pays the bribe in the teacher's class. Therefore, the payoff of the teacher depends on the number of students in the class n and of the number of students paying the bribe in the class k_b. In *fixed-wage* the payoff is

$$\pi_f(n, k_b) = F + b \cdot k_b$$

and in *piece-rate* is

$$\pi_{pr}(n, k_b) = s \cdot n + b \cdot k_b$$

When a teacher does not solicit bribes $k_b = 0$ and the payoffs π_f and π_{pr} change accordingly.

3.3 Equilibrium analysis

We make the following **assumptions** about the parameters of the game.

- Assumption 1: $\pi_{out} < 1 - b - a$. This implies that it is better to pay the bribe and obtain a diploma with a bad reputation than to drop out from school and take the outside option.
- Assumption 2: $\bar{c} > b + a$. This implies that some bad students are better off when they are in a school where all the students pay the bribe compared to when they obtain the diploma by studying hard.
- Assumption 3: $\mathcal{F}(a + b) > \frac{2}{3}$. This implies that, for a given size of the bribe, we look at cases where the reputation damage of bribes is not too low—i.e., there are enough students motivated to avoid such a reputation cost.
- Assumption 4: If the teachers are paid in a piece-rate regime, the piece-rate s satisfies $s > t \cdot b$, where
$$t = \max\left(\frac{1}{2\mathcal{F}(a+b) - 1}, \frac{2(1 - \mathcal{F}(a+b))}{3\mathcal{F}(a+b) - 2}\right).$$
A sufficiently large piece-rate makes bribe solicitation the dominated choice for the teachers.

We focus on the symmetric perfect Bayesian Nash equilibrium in which all the students use the same strategy when deciding to pay or not to pay a bribe. Given assumptions 1-4, the following proposition describes the strategies of the players in the unique symmetric perfect Bayesian equilibrium. The proof of the proposition is reported in Appendix B.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 $c > 1 - \pi_{out}$ do not go to school and students with $c \leq 1 - \pi_{out}$ 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 $c > b + a$ choose the teacher soliciting 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 students have chosen a teacher who solicits bribes, the students with $c > b + \frac{1}{n}a$ 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.*

The intuition 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 to attract some bad students, i.e., those students with $c > a + b$, 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 dominant for the 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 high 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 $c > a + b$, 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. This demand for bribe-free schools in combination with the assumption that the reputation damage of bribes is not too low (Assumption 3) implies the following: when both teacher are soliciting bribes, a teacher who stops soliciting bribes can attract the majority of the students, and hence obtain a better pay.

Assumption 4 points out that the piece-rate needs to be high enough to stop teachers from soliciting bribes. Without this assumption there are different equilibria depending on the level of s. If 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 equilibria and teachers are in a coordination game; (ii) one teacher solicits and the other does not solicit bribes is an equilibrium and teachers are in a "battle of the sexes" type of game.

4 Experimental design and hypotheses

4.1 Experimental design

We have two treatments, i.e., the fixed-wage (FW) treatment and the piece-rate (PR) treatment, that are based on the two institutional settings described in the previous section. The following parameters are common to both treatments: the diploma is worth at maximum 250 points, the reputation cost a is equal to 100 points, the bribe b is equal to 10 points, and the outside option's payoff π_{out} is equal to 115 points. As for the students' effort costs, students can be of one of the following types: Good ($c = 10$), Medium ($c = 65$), and Bad ($c = 160$), with probability $\frac{1}{6}$, $\frac{4}{6}$, and $\frac{1}{6}$, respectively.

Both treatments consist of two parts of 15 periods each, in which groups of 10 participants, 2 playing as teachers and 8 playing as students, repeatedly interact. The first part, i.e., the first 15 periods, is the same in both treatments. In this part teachers are paid a *low fixed-wage* of 40 points. Note that in this part teachers receive a fixed-wage that is lower than the payoff students can obtain. The rationale for this is to facilitate the emergence of a norm of corruption which is a prerequisite to test the effectiveness of the two interventions we analyze.

The second part, periods 16-30, introduces an intervention. We consider two different interventions aimed at breaking the established corrupt norm. The interventions vary by treatment in a between-subjects design. In the FW treatment teachers obtain a raise of 200 points and, hence, a *high fixed-wage* of 240 points. In the PR treatment, instead, teachers obtain a piece-rate bonus of $s = 50$ points for each student choosing their class on top of their fixed payment of 40 points. Note that the two interventions are neutral in terms of public expenditure. In both cases, the same amount of money is invested to reduce corruption, i.e., 400 points.

Given the Proposition, the experimental parameters induce the following equilibrium behavior. If both teachers do not solicit bribes, Bad students ($c = 160$) prefer not to go to school, while Medium ($c = 65$) and Good ($c = 10$) 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; Medium 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, Medium 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.

4.2 Hypotheses

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

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

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 (Section 3) and the case of teachers who are driven by other-regarding concerns.

4.2.1 Predictions for self-interested teachers

As for the case of self-interested teachers, note that the experimental set-up satisfies the assumption of the model. 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.

4.2.2 Predictions for teachers with other-regarding concerns

When one teacher refrains from soliciting bribes, all the students with $c \leq a+b$, choose to study with this teacher and, by doing so, avoid the reputation costs of bribing. Most of these students are therefore better off compared to when both teachers are soliciting bribes. Predictions whether this is enough to reduce bribe solicitation, however, crucially depends on the level of inequity aversion and on the actual reduction of inequality when teachers are not soliciting bribes. The latter depends on both the wage levels of the teachers after the pay-rise and on the distribution of students' effort costs. This renders the derivation of predictions for the general case quite cumbersome, we therefore restrict the predictions for the case of other regarding preferences to the constellation of parameters used in the experiment.

With these parameters, the majority of the students—i.e., the Medium and the Good students—are better off when a teacher does not solicit bribes compared to when both teachers are soliciting bribes. Since teachers are paid a high fixed-wage that is higher than the students payoffs, not soliciting bribes reduces inequality. A pay-rise in the fixed-wage setting can therefore break a corrupt system if teachers dislike inequality. To illustrate this we derive predictions assuming that teachers are inequity averse with Fehr and Schmidt (1999) preferences, i.e.,

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

where x_i is the payoff of the agent—i.e., the inequity averse teacher—, x_{-i} is the vector of payoffs of the other agents—i.e., the students and the other teacher—, α is the parameter measuring aversion to disadvantageous inequality, and β is the parameter measuring aversion to advantageous inequality. Appendix D reports the derivation of the predictions. Figure 1 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). Specifically, panel (a) shows the sets of equilibria for the parameters in the first part of both treatments, i.e., the low fixed-wage of 40, panel (b) shows the set of equilibria for the second part of the *FW* treatment; and panel (c) shows the set of equilibria for the second part of the *PR* treatment. The dots in each figure reports different estimated values of α and β reported in the literature (Goeree and Holt, 2000; Blanco et al., 2011; Beranek et al., 2015).

In the only equilibrium of the first part of the experiment both teachers solicit bribes for any reasonable value of the inequity aversion parameters since teachers earn considerably less than students. As for the second part, in *FW* teachers obtain a much higher fixed payoff that is as much as the good students earn when they obtain their diploma by studying hard. For moderate levels of aversion to advantageous inequality, the only equilibrium is to not solicit bribes, since otherwise the inequality towards the other teacher and most of the students is increased.

The effect of inequity aversion in the second part of the *PR* is less straightforward. For reasonable values of α and β, two equilibria emerge: one where teachers do not solicit bribes and the other where they solicit 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.

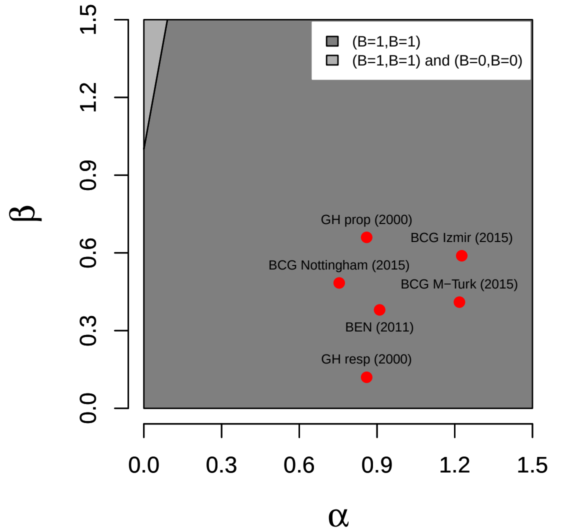
This analysis shows that a fair wage with reasonable levels of inequity aversion can break corruption in our experiment. Based on this, Hypothesis H3 will be rejected in favor of the alternative hypothesis that in treatment *FW*, the fraction of bribes solicited (and paid) significantly decreases in the second part compared to the first part.[8]

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 solicitation of bribes can be derived if teachers are "imperfect conditional cooperators" who match others' contributions only partly.

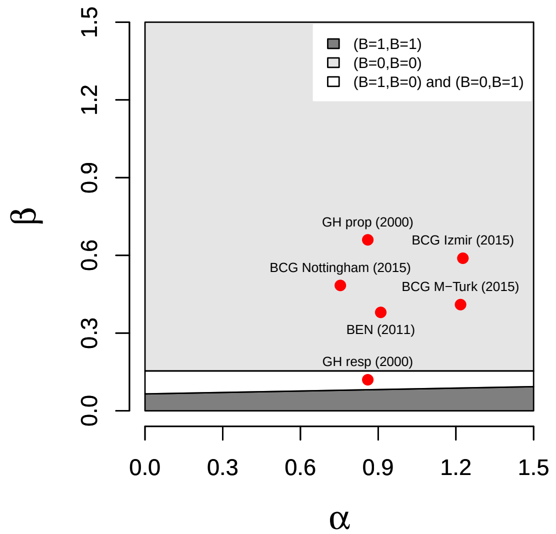
[8] In this analysis we continue to make the simplifying assumption that students are selfish. Students' choices have a lesser impact on the distribution of payoffs than teachers' choices. For a range of parameters that includes the estimates of Blanco et al. (2011), the predictions listed above remain the same when inequity aversion among students is introduced.

Figure 1: Equilibria in part 1 and part 2 of the two treatments when teachers are inequity averse with Fehr and 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 and 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.

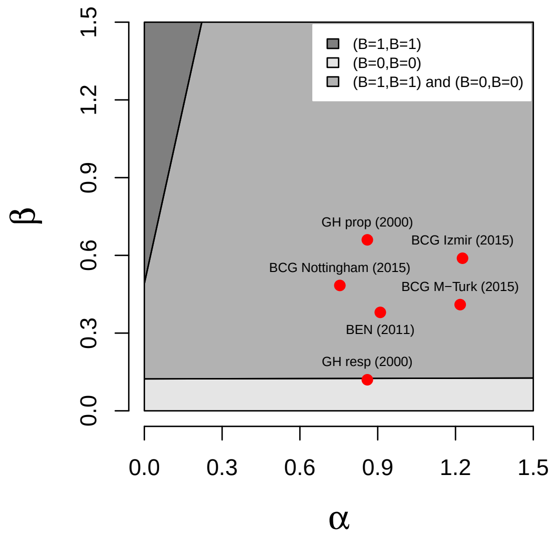
(a) FW and PR - Part 1



(b) FW - Part 2



(c) PR - Part 2



Fischbacher and 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.

4.3 Experimental procedures

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 because corruption and associated impunity are widespread in the education sector there (OECD, 2016). The sample's average age was 20.00 ($SD = 2.1$) and 57.7% of the participants were female. Participants were randomly assigned into a FW or a PR treatment. In each treatment they were randomly matched into 24 groups of 10 participants each.

When entering the laboratory, participants were randomly assigned to visually isolated computer terminals and were asked to read and sign a participation consent form.[9] Afterwards, 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 of 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.

In the experiment, we explicitly labeled participants' roles as teachers and students. In addition, we referred to the bribe as as a "motivation fee". The decision to use framed instructions was motivated by both 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 most of the experimental literature on corruption games (see, e.g., Abbink and Hennig-Schmidt, 2006; Armantier and Boly, 2013; Barr and Serra, 2009).

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, 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".

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 are higher than the salary a participant can earn in a comparable two hours students' job.

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

5 Results

5.1 Teachers' behavior and overall frequency of bribing

5.1.1 Bribes solicited by the teachers

The violin plots of Figure 2 illustrate the fraction of bribes solicited by the teachers. In part 1 (panel (a)), bribery levels exceed 90% for the majority of the groups and do not differ across the two treatments (Wilcoxon rank sum test $p = 0.426$).[10] 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.

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 with self-regarding preferences by not rejecting H1 & H3, while rejecting H2 & H4.

To corroborate these findings, we conducted a series of linear probability models that are estimated using individual level data, 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 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.[11]

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 ($Period\times$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*)$\times$d(*part 2*)) and (ii) this difference between treatments increases over periods ($Period\times$d(*piece-rate*) in Model 2 and $Period\times$d(*piece-rate*)$\times$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.

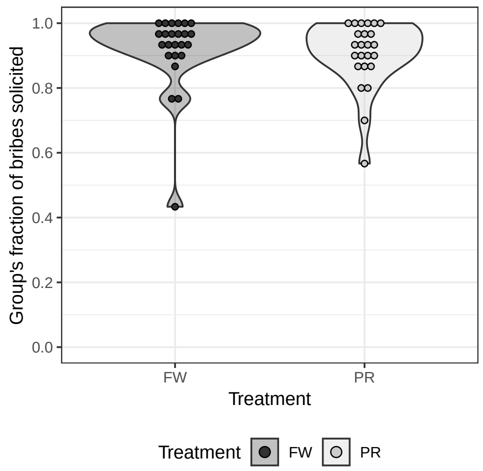
Now we turn to the dynamics in teachers' behavior in the second part of the experiment to gain deeper insights into how the shift from a corrupt system to a bribe-free system occurs. In particular, we study how teachers react to their own choices as well as other teachers' choices

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

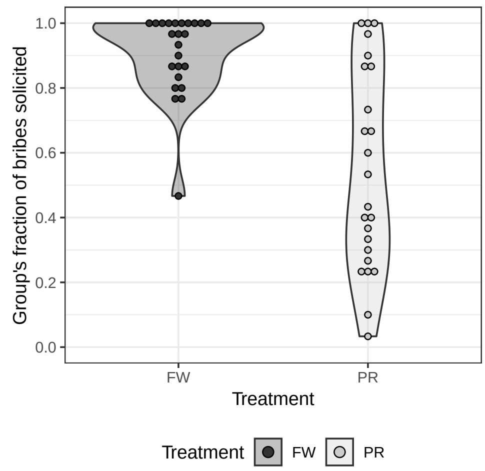
[11] Logit and Probit regressions with errors clustered at the group level provide similar results. The differences compared to the linear probability model are in the interaction term $Period\times$d(*piece-rate*)$\times$d(*part 2*) of Model 3 that becomes non-significant and in the control variable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

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

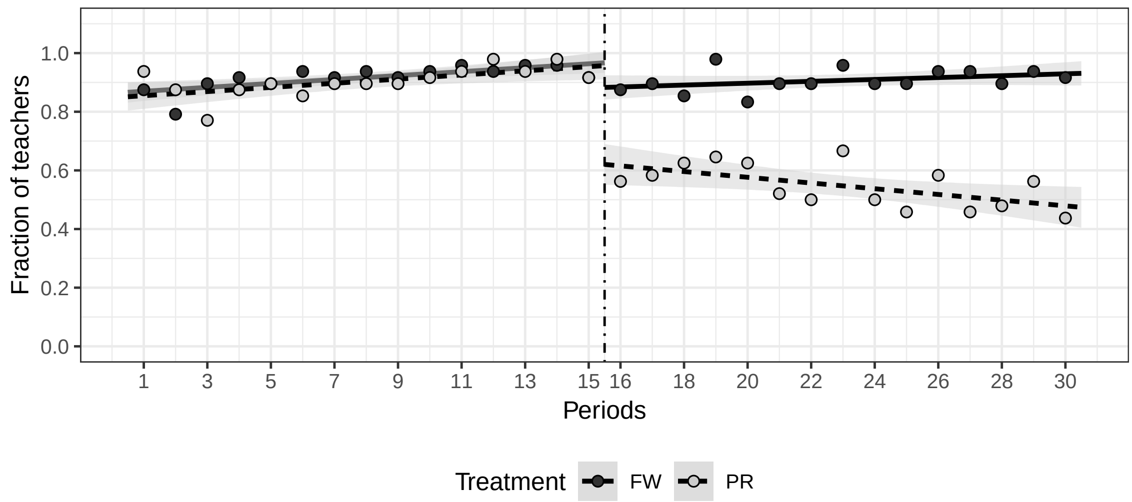
(a) Part 1



(b) Part 2



(c) Fraction of teachers soliciting bribes over periods



in previous rounds. Table 2 reports the fraction of times a teacher solicits bribes conditional on their choice and the choice of the other teacher in the previous period for the two treatments. In the FW treatment, teachers tend to solicit bribes independently 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 only if the other teacher did so in the previous period (more strongly if both did so) while they tend to abstain from soliciting bribes if the other teacher also abstained. This is not in line with the approach 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 like behavior that is compatible with the support that Fischbacher and Gächter (2010) document for incomplete conditional cooperation.

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: $0 < *** \leq 0.001 < ** \leq 0.01 < * \leq 0.05 < ^\circ \leq 0.10$

Table 2: 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	+	+	+	+

5.1.2 Overall frequency of bribing.

Besides analyzing the fraction of bribes solicited by the teachers, we additionally analyze the overall occurrence of bribery. Note that this 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 E.1 in appendix E shows the overall frequency of bribes paid in the group by part and by treatment and the fraction of bribes paid over periods.

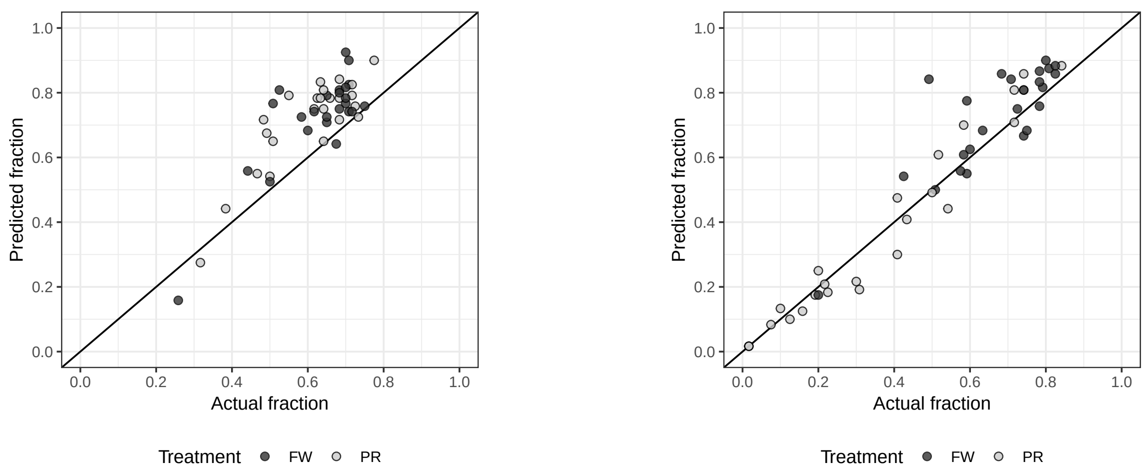
Columns 4-6 of Table 1 report a series of linear probability models that are estimated using individual level data. The dependent variable is a dummy taking value 1 if a student paid a bribe in that period. Explanatory variables are the same as in Table 1. Results confirm the observed effects of teachers' 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 exerted a significant positive effect on bribe payments, translating into a 7 percentage points increase.

5.2 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 3, 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 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)).

Figure 3: 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.

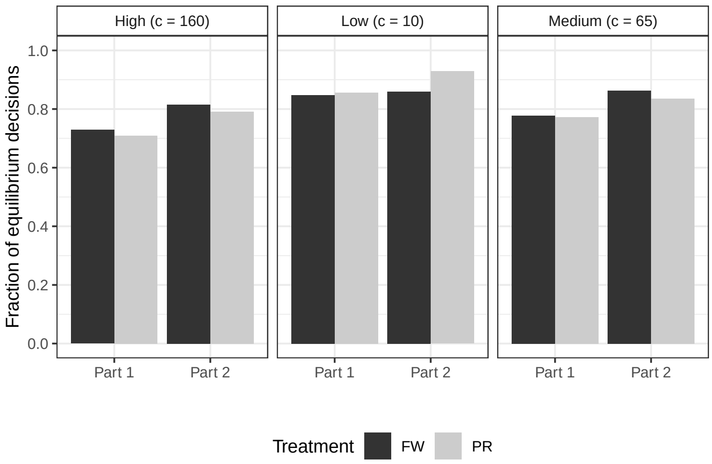
(a) Part 1 (b) Part 2



Furthermore, Figure 4 reports the fraction of choices in line with equilibrium predictions by effort cost and treatment. Results suggest that the likelihood to observe students playing the strategy predicted by the standard approach is generally high in both parts of the experiment (overall 78% in part 1 and 84% in part 2).

Alternative motives for the students' choices may exist. A prime candidate are moral concerns that curb students' willingness to pay bribes (Köbis et al., 2016). Additional analyses reported in the appendix E 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 provide very little support for the presence of such moral concerns in the current set-up.

Figure 4: Students' fraction of choices in line with the equilibrium strategy by effort cost and treatment

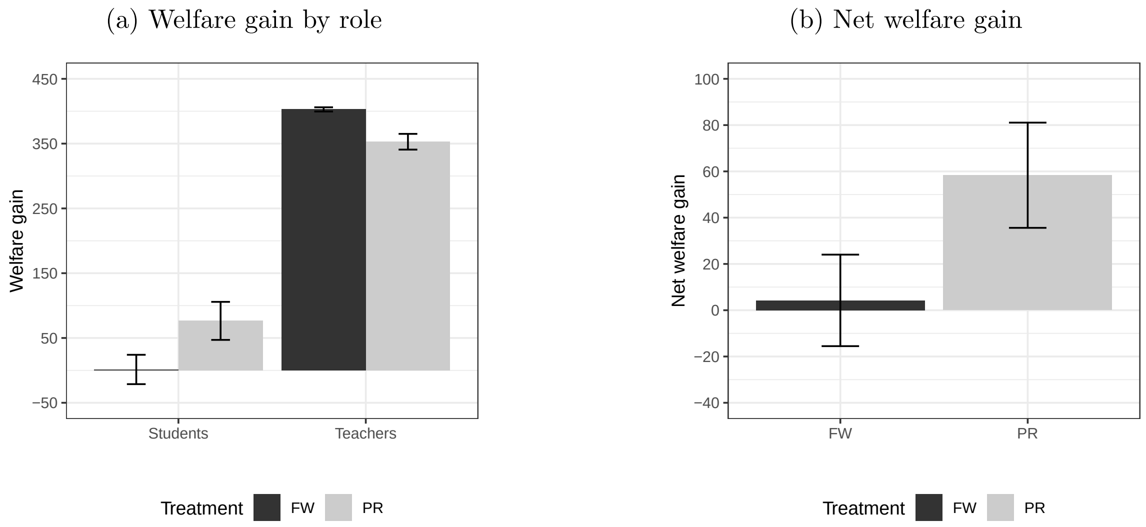


5.3 Social welfare

As a final step, 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 5 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).

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

Figure 5: 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.)



$p < 0.001$). 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 5 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).

6 Discussion

Can a corrupt education system be changed through a teachers' salary reform? Although being a frequently used anti-corruption policy, the theoretical predictions and behavioral evidence reveal that a substantial fixed-wage increase for teachers fails to break the previously established corrupt system. As an alternative anti-corruption policy, we introduce a piece-rate scheme that pays teachers according to the number of students in a school. By creating a market for integrity, this novel policy successfully reduces the occurrence of bribery transactions and increases overall social welfare. It represents a compelling counter-example to the widely held belief that competition fosters corruption.

Our lab experiment allows a controlled investigation into the dynamics of corruption in the education by using a relatively simple version of the game that captures the key strategic features of the relevant interaction. We regard its result as a proof of concept supporting the idea that a piece-rate regime may help fighting corruption in the education sector. 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 and Boly (2013) provide supportive evidence that lab and field data on corruptive practices in the education sector largely overlap. Moreover, we discuss several of additional variations of the model below. Using the assumption that the players are selfish, which is to a large extent supported by the experimental data, the predictions for the fixed-wage regime remain the same in the extensions that we consider here.[12] That is, teachers will continue to cash-in on higher wages without changing their corruptible 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 good as the fixed-wage regime. Appendix B present formal analyses of these points.

6.1 Extensions of the model

As a first extension let us consider what happens when the number of schools differs from two. Indeed, for students living in a rural area only one school might be available, whereas for students in densely populated areas more than two schools might exist. With one school, a piece-rate regime will 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 (see appendix B.3). Note that, even though corruption is reduced to a lesser extent in such a case, 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 "islands of integrity" see for example, Jackson and Köbis, 2018; Zuniga, 2018).

A second extension of the model considers the fact that teachers cannot handle an unlimited number of students and examines the implications of an upper-bound on the maximum class-size (see appendix B.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 the 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.

Third, in practice the subsidies offered in the piece-rate regime may encourage teachers to consider offering students a rebate when they join their school. Like the previous variations,

[12] As a short note on the lack of support for the social preference model, it is conceivable that inequity aversion failed to lower the bribe levels after the introduction of the fixed-wage increase because—in contrast to previous work estimating inequity aversion parameters—our set-up entailed multiple potential sources of inequity (i.e. multiple players). Moreover, the relevant payoff calculations are relatively complex, which may provide wiggle room to ignore payoff consequences for the others.

rebates may dampen the effectiveness of the piece-rate regime, but they will not neutralize it. With rebates, solicitation of bribes emerges (see appendix B.4). The reason lies in the competition for students that forces teachers to increase the rebates such that their profits vanish. Again, one can show that conditions exists under which, in equilibrium, one teacher will be targeting good students and not solicit bribes, while the other will be targeting bad students and solicit bribes.

A final extension of the model consists of allowing teachers to choose the size of the bribes that they solicit (see appendix B.2). In our base-line model, the size of the bribes is fixed. Teachers merely decide whether or not to solicit bribes. Although continuous bribes complicate the strategic decisions, they do not affect the main predictions. Akin to the base-line model, sufficient subsidies lead both teachers to abstain from soliciting bribes.[13]

6.2 Other considerations regarding external validity

One possible concern is that the introduction of incentives may erode an intrinsic motivation to perform well. For instance, Gneezy and Rustichini (2000a,b) find that the introduction of small monetary penalties in day-care centers for parents who arrive late to collect their children increases late-coming by the parents. Likewise, there is a danger that a monetary incentive reduces the motivation of well-performing teachers. We think that a piece-rate regime should primarily be considered where education is of poor quality due to lack of motivation. For example, high levels of absenteeism in many developing countries severely undermine the quality of education (Chaudhury et al., 2006; Rogers and Vegas, 2009). The proposed piece-rate regime may not only be valuable to fight corruption per se, but also contribute to an improvement in quality of education when students "punish" absent teachers by "voting with their feet".

Our model further abstracts from potential selection effects. Corruption and absenteeism may partly be explained by the possibility that meagre wages are only attractive to a specific group of teachers who do not mind to enhance their wages by a combination of corruptible behaviors and participation in side jobs. Both an increase in the fixed-wage and the introduction of a piece-rate regime may help attracting better motivated teachers. If there is a difference in the selection effect of the two mechanisms, we expect that a piece-rate regime will outperform the fixed-wage increase because in the latter under-performing incumbent personnel will fight harder against being replaced.

6.3 Challenges to implement a piece-rate regime

The proposed piece-rate scheme requires several preconditions in order for it be effective. 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.

[13]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 to pay bribes and, at the same time, gives students a stronger incentive to avoid bribing classrooms. Other than that, the theoretical predictions of the base-line model remain the same.

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 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 and 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; UNDP, 2011).

Third, piece-rate based interventions requires calibration of the incentives. Our model predicts that a piece-rate regime will be successful only if the subsidy is sufficiently high to trigger actual competition for students. 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 and 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.

7 Conclusions

Corruption represents a major obstacle for the attainment of high quality education. Vast financial resources have been invested into improving education and alleviating corruption in education systems—especially in developing nations. One popular approach to counteract the negative effects of corruption are 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 corruption—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 corruption-stricken education systems, we propose a payment scheme 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 marks an extension to the existing 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 this piece-rate regime—although not eliminating corruption altogether, it substantially reduced its occurrence. The effectiveness of the piece-rate intervention bears additional weight when considering that it was introduced after a culture 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.

The strong results of our experiment bolster our conjecture that compared with a fixed-wage increase, enhancing wages through a piece-rate regime may be the more promising avenue to reduce corruption in education. Our data suggest that during the recent teacher salary reform in Indonesia (De Ree et al., 2015), a great opportunity to implement and test a promising alternative payment scheme was wasted. In our view, the next developing country that wants to enhance low wages in the education sector should seriously consider and pilot a piece-rate regime. If the piece-rate regime is introduced at varying dates as the fixed-wage increase was in Indonesia, the effectiveness of a piece-rate regime in the field can be assessed.

References

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

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

Abdulkadiroğlu, A., Pathak, P. A., and Walters, C. R. (2018). Free to Choose: Can School Choice Reduce Student Achievement? *American Economic Journal: Applied Economics*, 10(1):175–206.

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

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

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

Armantier, O. and 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.

Azfar, O. and Nelson, W. R. (2007). Transparency, wages, and the separation of powers: An experimental analysis of corruption. *Public Choice*, 130(3-4):471–493.

Baldacci, E., Clements, B., Gupta, S., and Cui, Q. (2008). Social Spending, Human Capital, and Growth in Developing Countries. *World Development*, 36(8):1317–1341.

Banerjee, A. V., Banerji, R., Duflo, E., Glennerster, R., and Khemani, S. (2008). *Pitfalls of participatory programs: Evidence from a randomized evaluation in education in India.* The World Bank.

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

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

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

Bicchieri, C. (2016). *Norms in the wild: How to diagnose, measure, and change social norms.* Oxford University Press.

Blanco, M., Engelmann, D., and Normann, H.-T. (2011). A within-subject analysis of other-regarding preferences. *Games and Economic Behavior*, 72(2):321–338.

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

Cadot, O. (1987). Corruption as a Gamble. *Journal of Public Economics.*

Cappelen, A. W., Fjeldstad, O.-H., Mmari, D., Sjursen, I. H., and Tungodden, B. (2018). Understanding the resource curse: A large-scale experiment on corruption in tanzania. *CMI Working Paper.*

Cappelen, A. W. and Tungodden, B., editors (2019). *The Economics of Fairness.* Elgar research collection. Edward Elgar Publishing.

Chaudhury, N., Hammer, J., Kremer, M., Muralidharan, K., and 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., and 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., and Urzua, S. (2010). The education-health gradient. *American Economic Review*, 100(2):234–38.

Cooray, A. and 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., and Rogers, H. (2015). Double for nothing? the effect of unconditional teachers' salary increases on performance. *Quarterly Journal of Economics.*

Denisova-Schmidt, E. and Prytula, Y. (2017). Ukraine: Endemic higher education corruption. *International Higher Education*, 90:16–18.

Di Tella, R. and 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.

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

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

Falk, A. and Heckman, J. J. (2009). Lab experiments are a major source of knowledge in the social sciences. *Science*, 326(5952):535–538.

Falk, A. and Szech, N. (2013). Morals and markets. *science*, 340(6133):707–711.

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

Ferraz, C., Finan, F., and Moreira, D. B. (2012). Corrupting learning: Evidence from missing federal education funds in brazil. *Journal of Public Economics*, 96(9-10):712–726.

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

Fischbacher, U. and 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. and Golden, M. A. (2017). *Corruption: What everyone needs to know*. Oxford University Press.

Foltz, J. D. and 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.

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

Gneezy, U. and Rustichini, A. (2000a). A fine is a price. *The Journal of Legal Studies*, 29(1):1–17.

Gneezy, U. and Rustichini, A. (2000b). Pay enough or don't pay at all. *The Quarterly Journal of Economics*, 115(3):791–810.

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

Hallak, J. and 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., and Nuraliyeva, N. (2008). The Cost of Corruption in Higher Education. *Comparative Education Review*, 52(1):1–25.

Heywood, P. M. (2018). Combating corruption in the twenty-first century: new approaches. *Daedalus*, 147(3):83–97.

Hoxby, C. M. (1994). Do private schools provide competition for public schools? Technical report, National Bureau of Economic Research.

Hoxby, C. M. (2000). Does competition among public schools benefit students and taxpayers? *American Economic Review*, 90(5):1209–1238.

Hsieh, C.-T. and Urquiola, M. (2006). The effects of generalized school choice on achievement and stratification: Evidence from Chile's voucher program. *Journal of Public Economics*, 90(8-9):1477–1503.

Jackson, D. and Köbis, N. C. (2018). *Anti-corruption through a social norms lens.* U4 - Anti Corruption Resource Center.

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., and Van Lange, P. A. (2016). Prospection in individual and interpersonal corruption dilemmas. *Review of General Psychology*, 20(1):71.

Lambsdorff, J. G. and 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.

Mishra, P., Subramanian, A., and Topalova, P. (2008). Tariffs, enforcement, and customs evasion: Evidence from India. *Journal of Public Economics*, 92(10-11):1907–1925.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., Group, P., et al. (2009). Preferred reporting items for systematic reviews and meta-analyses: the prisma statement. *PLoS medicine*, 6(7):e1000097.

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

Nambissan, G. B. and 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.

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

Olken, B. A. (2009). Corruption perceptions vs. corruption reality. *Journal of Public economics*, 93(7-8):950–964.

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

Rogers, H. F. and Vegas, E. (2009). *No more cutting class? Reducing teacher absence and providing incentives for performance.* The World Bank.

Rose-Ackerman, S. and 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. and 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.

Shleifer, A. and Vishny, R. W. (1993). Corruption. *The quarterly journal of economics*, 108(3):599–617.

Sutter, M., Haigner, S., and 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.

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.

TI, T. I. (2005). *Stealing the Future – Corruption in the Classroom*. Transparency International.

TI, T. I. (2013). *Global Corruption Barometer*. Transparency International.

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

UNDP, U. N. D. P. (2011). *Fighting Corruption in the Education Sector. Methods, Tools and Good Practices*. United Nations Development Programme.

Van Rijckeghem, C. and 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.

Verger, A., BonaL, X., and Zancajo, A. (2016). What are the role and impact of public-private partnerships in education? a realist evaluation of the chilean education quasi-market. *Comparative Education Review*, 60(2):223–248.

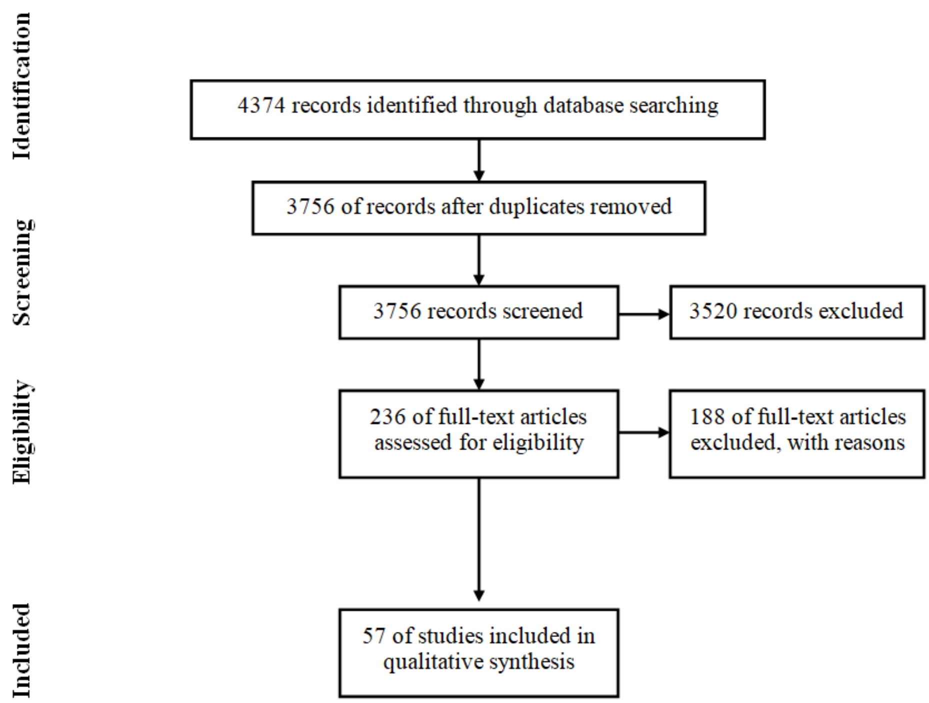
Zuniga, N. (2018). *The impact of 'islands of integrity'*. U4 - Anti Corruption Resource Center.

A Systematic Literature Review. [ONLINE APPENDIX]

A.1 Search for studies.

We systematically searched the databases of Web of Science, Jstor, PiCarta, Scopus and Google Scholar using the following keywords and Boolean operators: "Fair salary" OR "fair wage" OR "salary increase" OR "salary top-up" OR "wage increase" AND corruption OR bribery OR bribe OR rent-seeking OR shirking OR fraud OR embezzlement OR "academic cheating" We additionally searched for all articles that cited highly relevant papers on this topic (i.e. Abbink, 2002; Akerlof and Yellen, 1990; Di Tella and Schargrodsky, 2003; Van Rijckeghem and Weder, 2001; Foltz and Opoku-Agyemang, 2015). Furthermore, we issued a call for papers through several mailing lists in order to identify all relevant unpublished papers. The search was performed in January and February 2018. After removing duplicates, 3764 articles were identified (for a detailed illustration see the PRISMA chart in Figure A.6; Moher et al., 2009). One independent coder inspected all entries to see whether they meet the pre-specified inclusion criteria (see more details on Open Science Framework).

Figure A.6: PRISMA chart



A.2 Results - Systematic Literature Review

In total, 57 studies are included in the review. Of these, 17 show an inverse relationship (i.e., an increase in salary leads to a decrease in corruption), 16 studies show no relationship, 7 articles a positive association (i.e., an increase in salary is associated with an increase in corruption) and 16 studies report mixed results on the effect of salary on corruption. One study concludes a quadratic relationship. The corruption and salary measures as well as the effects are summarized in Table A.3.

Table A.3: Papers included in the systematic literature review.

Study No	Study	Total N	Country	Effect of fixed-wage increase
1	Abbink (2000)	24 (P)	Germany	0
2	Alt & Lassen (2012)	1114 (O)	USA	0
3	Amin (2015)	13 (O)	Indonesia	0
4	An & Kweon (2017)	129 (C)	129 countries	-
5	Armantier & Boly (2008)	138 (P)	Burkina Faso, Canada	-/0/+
6	Armantier & Boly (2011)	76 (P)	Burkina Faso	-/0/+
7	Armantier & Boly (2013)	202 (P)	Burkina Faso, Canada	-/0/+
8	Azfar & Nelson (2007)	96 (P)	USA	-
9	Barr et al. (2009)	144 (P)	Ethiopia	-
10	Benito et al. (2015)	110 (O)	Spain	-
11	Benito et al. (2018)	358 (O)	Spain	+
12	Borcan et al. (2014)	2459 (O)	Romania	-
13	Braendle (2015)	1319 (P)	across 27 countries	-/+
14	Chen & Sandino (2012)	76 (O) & 133 (O)	USA	-
15	Dabalen & Wane (2008)	1278 (P)	Tajikistan	-/0
16	De la Torre et al. (2017)	384 (P)	Austria	0
17	de Ree et al. (2016)	360 (O)	Indonesia	0
18	Di Tella & Schargrodsky (2003)	544 (P)	Argentina	-/0
19	Dong & Torgler (2013)	31 (O)	China	-
20	Eitle et al. (2014)	497 (O)	USA	0
21	Ferraz & Finan (2008)	Min. 4010 (O)	Brazil	-/0
22	Flory et al. (2017)	81 (P)	USA	-
23	Foltz & Opoku-Agyemang (2015)	2400 (O)	Burkina Faso, Ghana	+
24	Frank & Schulze (2000)	190 (P)	Germany	0
25	Gnimassoun & Keneck (2015)	130 (C)	130 countries	0
26	Gong & Wu (2012)	Unknown	China	+
27	Greenberg (1990)	143 (P)	USA	-
28	Gurgur & Shah (2005)	30 (C)	across 32 countries	0
29	Herzfeld & Weiss (2003)	40-72 (C)	across 130 countries	0/+

30	Jacquement (2005)	87 (P)	France	-/+
31	Johnson (2012)	384 (O)	USA	0/+
32	Karahan et al. (2006)	82 (O)	USA	+
33	Kim, Bong Hwan (2015).	114 (P)	Cambodia	-
34	La Porta et al. (1999)	63 (O)	across 204 countries	+
35	Le et al. (2013)	930 (O)	across 73 countries	-
36	Lindkvist (2014)	145 (P)	Tanzania	0/-
37	Mahmood (2005)	67 (C)	across 67 countries	-
38	Matuszewski (2010)	224 (P)	USA	-/0/+
39	Mocan & Altindag (2011)	1151 (P)	across 27 countries	+
40	Navot et al. (2016)	18.800 (O)	across 53 countries	+
41	Panizza et al. (2001)	27 (O)	across 17 countries	0
42	Pellegrini (2011)	73 (C)	across 73 countries	-/0
43	Peter & Zelenska (2010)	20368 (P)	Russia	0/+
44	Rauch & Evans (2000)	35 (C)	across 35 countries	0
45	Rubin & Whitford (2008)	34 (C)	across 34 countries	0
46	Sardzoska & Tang (2014)	515 (P)	Macedonia	0
47	Schulze & Frank (2003)	390 (P)	Germany	-/0
48	Schulze et al. (2016)	79 (O)	Russia	-
49	Subramanian & Chakrabarti (2011)	146 (O)	across 26 countries	U-shaped
50	Tang (2003)	211 (P)	China	0
51	Tang, T. et al., (2002)	2338 (P)	across 12 countries	0
52	Treisman (2000)	47 (C)	across 99 countries	0
53	Van Rijckeghem & Weder (1997)	25 (C)	across 22 countries	-
54	Van Rijckeghem & Weder (2001)	31 (C)	Unknown	-
55	Van Veldhuizen (2013)	76 (P)	The Netherlands	-
56	Vegas & De Laat (2003)	2300 (P)	Togo	-
57	Voors et al. (2011)	569 (O)	across 29 countries	-/+

Notes: The table contains studies conducted on the individual, organizational and country level. Annotations in column Total N refer to these different level of analysis by (P) referring to number of participants in the sample; (O) referring to number of organizations in the sample; (C) referring to number of countries in the sample; The observed effect of a fixed-wage increase is displayed in the last column, with - referring to a decrease of corruption; 0 referring to a null effect, and + referring to a increase of corruption

B Theory: base model and extensions

B.1 Proof of the main proposition

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

Stage 3: choice whether or not to pay the bribe. We first consider the shortest subgame in which a student entered a school where the teacher solicit bribes. It is profitable for the student to pay the bribe if $\Pi(1, n_b, n, c) > \Pi(0, n_b, n, c)$. That is the case when the effort saved by bribing is bigger than the bribe and the damage of reputation produced by the student, i.e., $c > b + \frac{1}{n}a$, or when

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

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

Stage 2: choice of the school. We now consider the subgames in which students choose which school to attend, if any. Three subgames 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 $1 - c$ or (ii) to not go to school and get a payoff of π_{out}. Therefore, students with $c > 1 - \pi_{out}$ prefer not to go to school and students with $c \leq 1 - \pi_{out}$ prefer to go to school. Since the students who prefer to go to school are indifferent between the two teachers, we assume that they are randomly split between them.

- **One teacher solicit bribes and the other teacher does not.** All students use the same cutoff strategy, which means that all students with a c 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. c^* must be equal to $b + a$. Notice that given that other students are using this cutoff strategy, a bad student with $c > b + a$ prefers the teacher soliciting the bribe. He/she anticipates that all students who choose this school will choose to bribe given that for all of them $c > b + a > b + \frac{1}{N}a$. Thus, by choosing the teacher who solicits bribes, the student will receive a payoff of $1 - b - a$ which exceeds the payoff of the other teacher if $1 - b - a > 1 - c$, or when $c > b + a$. Moreover, even bad students prefer to pay the bribe over not going to school because of the assumption that $\pi_{out} < 1 - b - a$.

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

  Likewise, it cannot happen in equilibrium that that all students use a c^* that is smaller than $b + a$. To see this, note that for any $b + \frac{1}{k}a \leq c^* < b + \frac{1}{k-1}a$, where $k \in \{2, \ldots, N\}$, the best type who chooses the teacher who solicits bribes ($c = c^*$) chooses to bribe when the size of his school is k or larger, but not when it is smaller than k. For the school sizes where he/she does not bribe, he/she would have been better off by choosing the other teacher because the probability that one of the other students bribes is not 0. For the school sizes where he/she does bribe, he/she will be in a school where everyone bribes (because the others have at least as high effort costs), and he/she will earn $1 - b - a$, which is less than he would have received by choosing the other teacher. Therefore, for any $b + \frac{1}{N}a \leq c^* < b + a$, the best type who 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 $c < b + \frac{1}{N}a$ 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}. Given that students can obtain at least $1 - b - a$ by going to school and paying the bribe, they all prefer to go to school because $\pi_{out} < 1 - b - a$. For this case we assume that students are randomly split between the two teachers.

Stage 1: choice of the teachers. Now we consider the longest subgame in which both teachers simultaneously decide whether or not to solicit bribes. For teachers' decisions, the institutional setting needs to be taken into account, i.e., *fixed-wage* or *piece-rate*.

Teachers are paid *piece-rate*. The payoff of the teachers is 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 solicit bribes, they have an expected attendance of $\frac{N}{2}\mathcal{F}(1-\pi_{out})$ students each and hence 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 $N\mathcal{F}(a + b)$. 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 that solicits bribes and $sN\mathcal{F}(a + b)$ for the one that does not solicit bribes.

- If both teachers solicits 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 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 bribing in the class is $E_{bn} = nP_n$. Moreover, given that students are randomly split between teachers, the unconditional expected number of students bribing in the class 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, the payoff of the two teachers is $s\frac{N}{2} + bE_b$.

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

		T_2 Bribe	T_2 No-Bribe
T_1	Bribe	$s\frac{N}{2} + bE_b$	$(s+b)N(1-\mathcal{F}(a+b))$
	No-Bribe	$sN\mathcal{F}(a+b)$	$s\frac{N}{2}\mathcal{F}(1-\pi_{out})$

No-Bribe is a dominant strategy if this strategy is preferred both when the other teacher plays Bribe, i.e.,

$$sN\mathcal{F}(a+b) > s\frac{N}{2} + bE_b \tag{1}$$

and when the other teacher plays No-Bribe, i.e.,

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

Consider condition (1) and note first that $E_b \leq \frac{N}{2}$ because $\mathcal{F}\left(\frac{a}{n}+b\right) \geq 0$ which implies that $P_n \leq 1$. Therefore, $sN\mathcal{F}(a+b) > s\frac{N}{2} + b\frac{N}{2}$ is a stronger condition from which we obtain that

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

Consider condition (2) and recall that $\pi_{out} < 1-b-a$. Then $\mathcal{F}(1-\pi_{out}) > \mathcal{F}(a+b)$ and, hence, $s\frac{N}{2}\mathcal{F}(a+b) > (s+b)N(1-\mathcal{F}(a+b))$ is a stronger condition from which, assuming that the fraction of good students is $\mathcal{F}(a+b) > \frac{2}{3}$ of the population, we obtain

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

Therefore, a sufficient condition for No-Bribe to be dominant is that the fraction of good students in the population is $\mathcal{F}(a+b) > \frac{2}{3}$ and that the piece-rate s is at least t times the bribe where

$$t = \max\left(\frac{1}{2\mathcal{F}(a+b)-1}, \frac{2(1-\mathcal{F}(a+b))}{3\mathcal{F}(a+b)-2}\right)$$

This is reflected in Assumption 4. If these conditions hold, No-Bribe is the dominant strategy while teachers would be both better off cooperating and soliciting bribes.

Teachers are paid *fixed-wage*. The payoff of the teachers is a fixed amount F and, if they solicit bribes, they can obtain more money collecting the bribes paid.

This institutional setting is equivalent to the *piece-rate* setting where $s = 0$. In this case the payoff matrix for teacher T_1 is

		T_2	
		Bribe	No-Bribe
T_1	Bribe	$F + bE_b$	$F + bN(1 - \mathcal{F}(a+b))$
	No-Bribe	F	F

From the matrix we can see that, in case of *fixed-wage*, offering the bribe is a dominant strategy for the teachers. $\square$

B.2 Continuous bribes in the base model

In the base model we fixed the size of the bribe in case a teacher decides to solicit bribes. Here, we relax this assumption and assume that a teacher who chooses to be corruptible simultaneously announces the size of the bribe which may be any non-negative number. We show that both teachers not soliciting bribes remains an equilibrium in the piece-rate regime. When both teachers do not solicit bribes their payoff is

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

We assume

$$\pi_{out} \leq 1-a$$

which means that it is preferable to get a bad diploma for free than the outside option. We further assume that

$$b \leq \bar{c} - \frac{1}{N}a$$

which means that teachers do not set a bribe that no-one wants to pay in the biggest possible class. Finally, we assume that, when indifferent, students avoid strategic risk: i.e., students prefer to be in an non-bribing class compared to a class where bribes can be paid but are never paid.

Now consider a teacher who deviates to soliciting a bribe b while the other is not soliciting bribes.

[Case large bribes: $1-a-b \leq \pi_{out}$]. Consider first b such that $1-a-b \leq \pi_{out}$. We show that for such bribe values students do not join the bribing class.

1. Suppose that there are students joining the class and that all the students who join prefer to pay the bribe for all sizes of the class. In this case the students believe that only students with $c > a+b$ join the bribing class but then they have a profitable deviation to stay out: $1-a-b \leq \pi_{out}$.

2. Suppose that students prefer to join the bribing class and pay the bribe for some sizes of the class but not for others. Note that it must be that students join the bribing class for some $c \leq b+a$. If students join only for $c > b+a$ all the students pay the bribe for all sizes of the class and this contradicts the assumption. Moreover, there are no students joining the bribing class when $c \leq b + \frac{1}{N}a$. These students will never pay the bribe and, given that we assume that there are some students bribing for some size of the class, they have an incentive to deviate to avoid the externality. Therefore, students must join the bribing class for some c in the interval $\left(b+\frac{1}{N}a, b+a\right]$.

   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 $k \in \{1, 2, \ldots, N-1\}$ and take k^* such that there is a c for which students join the bribing class in the sub-interval $I(k^*)$ and there are no c for which students join the bribing class in all the sub-intervals $I(k)$ where $k > k^*$. Let c^* be a c 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 he pays the bribe all the other students in the class pay the bribe as well. Therefore his payoff is $1 - a - b$ when paying the bribe and at best $1 - c^*$ when not paying the bribe. The fact that $c^* < b + a$ implies that a student with such a c has a profitable deviation to join the teacher that is not soliciting bribes.

3. Therefore, it must be that all the students who join the bribing class prefer to never pay the bribe. In this case the students believe that only students with $c \leq 1 - \pi_{out}$ and $c \leq b + \frac{1}{N}a$ join the class otherwise they have a profitable deviation. These students, however, are equally well-off when joining the bribing and when joining the non-bribing class. In this case we assume that they do not join the bribing class in order to avoid strategic risk.

[Case small bribes: $1 - a - b > \pi_{out}$**].** If b is such that $1 - a - b > \pi_{out}$ we are in the same case as the one discussed in the base model with a fixed bribe. The payoff of the teacher soliciting bribes is

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

and in order not to have profitable deviations it must be that

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

for all $b \in [0, 1 - \pi_{out} - a)$. Note that: (i) $F(a+b)$ attains its minimum at $F(a)$, (ii) $\mathcal{F}(1 - \pi_{out}) \geq \mathcal{F}(a)$, and (iii) $b < 1 - a - \pi_{out}$. Therefore sufficient conditions to prevent profitable deviations are that the probability to have "bad" students is below $\frac{1}{3}$

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

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, this condition is satisfied when using the experimental parameters, i.e., $s = 50$, $a = 100$, $\pi_{out} = 115$, $\mathcal{F}(a) = \frac{5}{6}$. Moreover, it is possible to show that with these parameters, there are no equilibria where both teachers solicit bribes. This requires long and tedious calculations that are not included in appendix but are available from the authors upon request.

B.3 More than 2 schools in the base model.

In this section we analyze the case where there are $K > 2$ schools competing for students. Given the assumptions we make in the base model, there is a bribe-free equilibrium in the piece-rate regime with two schools, i.e., $s\frac{N}{2}\mathcal{F}(1-\pi_{out}) > (s+b)N(1-\mathcal{F}(a+b))$. This implies that

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

Suppose now 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 it is the case that

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

we prove that given K there is an equilibrium where some of the schools 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 are soliciting bribes and $K - Q^*$ teachers are not soliciting 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 not soliciting 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 soliciting 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 of the teachers soliciting bribes will be smaller than $s\frac{N}{K} + bE_b < s\frac{N}{2} + bE_b$.

B.4 Rebates in the base model.

Consider teachers providing monetary incentives to the students for joining the class. In this case one can show that there are no equilibria where both teachers do not solicit bribes if, as we assumed, there are students that prefer to obtain a diploma with a very bad reputation for free than to obtain a diploma exerting effort, i.e., $1-\bar{c} < 1-a-b$. Moreover, we can show that there may exist a Nash equilibrium where one teacher solicit and the other teacher do not solicit bribes depending on the models' parameter.

B.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 $c \leq 1-\pi_{out}+R_1$ go to teacher T_1 and the other students choose the outside option. In this case T_1 makes a negative payoff so $R_1 = s$ is a profitable deviation for T_1. If $R_1 \leq s$ all the students with $c \leq 1-\pi_{out}+R_1$ go to teacher T_1 and the other students choose the outside option. In this case T_2 does not attract students and he can profitably deviate to $R_2 = R_1$ where he obtains a payoff of $(s-R_2)\frac{N}{2}\mathcal{F}(1-\pi_{out}+R_2)$.

Assume now $R_2 = R_1$. If $R_1 = R_2 > s$ then T_1 makes a negative profit and he is better off deviating to $R_1 = s$. If $R_1 = R_2 < s$ teachers share the students that have an effort cost $c \leq 1-\pi_{out}+R_1$ and T_1 has a profitable deviation by setting $R_1 = R_1+\epsilon < s$. In this case 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 $c \leq 1-\pi_{out}+s$ and make a profit of 0. In this case, a teacher can accept bribes and offer the rebate $R = s$. This teacher will have a positive expected payoff because he/she will attract the students with $c \in (b+a, \bar{c}]$ and obtain $bN(1-\mathcal{F}(b+a))$.

B.4.2 One teachers solicit bribes and the other teacher do not solicit bribes

Suppose now 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 the students with an effort cost higher than c^* study with T_1 and pay the bribe for all sizes of the class and the students with an effort cost lower or equal than c^* study with T_2.

Let first look at the students. It must be the case that for students with $c > c^*$

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

and for students with $c \leq c^*$

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

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

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 $R_2^*(R_1)$ to look for possible equilibria. 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 both equation 3 and equation 4 are satisfied and $R_1^* < R_2^*$ a Nash equilibrium where one teacher solicit bribes and the other does not can be sustained, for instance, by the belief that students do not join the class of a teacher that switches from soliciting to not soliciting bribes (and vice-versa).

Finally, we provide a simple example showing that there are conditions under which such equilibrium exists. Assume that the student's effort cost is uniformly distributed between 0 and $\bar{c}$, then the payoff function of both teacher is a quadratic function in the rebate with support $[0, s + b]$ for T_1 and $[0, s]$ for T_2. Therefore the maximum payoff of the teachers 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 solutions where the teachers offer the positive rebate $R_1^* = s + \frac{1}{3}(a + 3b - 2\bar{c})$ and $R_2^* = s - \frac{1}{3}(a + \bar{c})$ and these rebates satisfies the students constraints. The following set of parameters $a = 0.2$, $b = 0.04$, $s = 0.3$, and $\bar{c} = 0.55$, for instance, gives $R_1^* = 0.04$ and $R_2^* = 0.05$ which satisfy the students' constraints that $R_1 \leq R_2$.

B.5 Effect of a cap on the size of the school with the experimental parameters.

In this section we discuss the effect of a cap in the size of the schools to k students. We do this exercise using the experimental parameters. Additionally, we assume that the total capacity of the two school is 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 behavior of the students is the following:

- When both teachers solicit bribes $(B = 1, B = 1)$, 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 $(B = 0, B = 0)$, the medium and good students decide to go to school and the bad students decide not to go to school. The students that decide to go to school are shared between the teachers. If n_B is the number of bad students, one teacher gets $\left\lceil \frac{N-n_B}{2} \right\rceil$ students and the other teacher gets $\left\lfloor \frac{N-n_B}{2} \right\rfloor$ students.

  Note that in both the previous cases the cap on the size is not binding and the payoffs of the teachers remain the same as in the non-capped case discussed in the main text.
- When one teacher solicit bribes and the other does not, i.e. either $(B = 1, B = 0)$ or $(B = 0, B = 1)$, the students apply to one of the school and, if there are too many students applying to one school, a random selection decides which one are the students in excess and they can decide either to go to the other school or not to go to school. Note that with this procedure the following is an equilibrium: (i) the good and medium students apply to the bribe-free class and, if rejected, they go to the bribing class; (ii) the bad students apply to the bribing class and, if rejected, they decide not to go to school. Therefore, letting n_B be the number of bad students and n_G be the number of good students, the payoff of the teacher not soliciting bribes conditional on n_G and n_B is

$$\Pi_1(1,0|n_B,n_G) = \begin{cases} 50 \cdot (N - n_B) & \text{if } N - n_B \leq k \\ 50 \cdot k & \text{if } N - n_B > k \end{cases}$$

  and the payoff of the teacher soliciting bribes conditional on n_G and n_B is

$$\Pi_1(0,1|n_B,n_G) = \begin{cases} 50 \cdot n_B + 10 \cdot n_B & \text{if } n_B \leq k \text{ and } N - n_B \leq k \\ 50 \cdot (N - k) + 10 \cdot (n_B + E_b) & \text{if } n_B \leq k \text{ and } N - n_B > k \\ 50 \cdot k + 10 \cdot k & \text{if } n_B > k \end{cases}$$

  where E_b is the expected number of medium students rejected from the other class when there are $N - n_B - n_G$ medium and n_G good students, i.e., the mean of an Hyper-geometric 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 the expected payoff of a teacher soliciting bribes is:

$$E\Pi_1(1,1) = 50 \cdot \frac{N}{2} + 10 \cdot \frac{N}{2} \cdot \frac{5}{6}$$

$$E\Pi_1(0,1) = \sum_{n_G=0}^{N} \sum_{n_B=0}^{N-n_G} \frac{N!}{n_G! n_B! (N - n_G - n_B)!} \left(\frac{1}{6}\right)^{n_G+n_B} \left(\frac{4}{6}\right)^{N-n_G-n_B} \Pi_1(0,1|n_B, n_G)$$

$$E\Pi_1(1,0) = \sum_{n_G=0}^{N} \sum_{n_B=0}^{N-n_G} \frac{N!}{n_G! n_B! (N - n_G - n_B)!} \left(\frac{1}{6}\right)^{n_G+n_B} \left(\frac{4}{6}\right)^{N-n_G-n_B} \Pi_1(1,0|n_B, n_G)$$

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

Table B.4 summarizes the effect of a cap of k on the size of the class when $N = 8$. When the cap is relatively big, 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 and undermines the incentive not to solicit bribes in piece-rate.

Table B.4: 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	$(B = 1, B = 1)$	$(B = 0, B = 0)$	$(B = 1, B = 0)$ and $(B = 0, B = 1)$
8		X	
7		X	
6		X	
5			X
4	X		

C 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@uva.nl)

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.

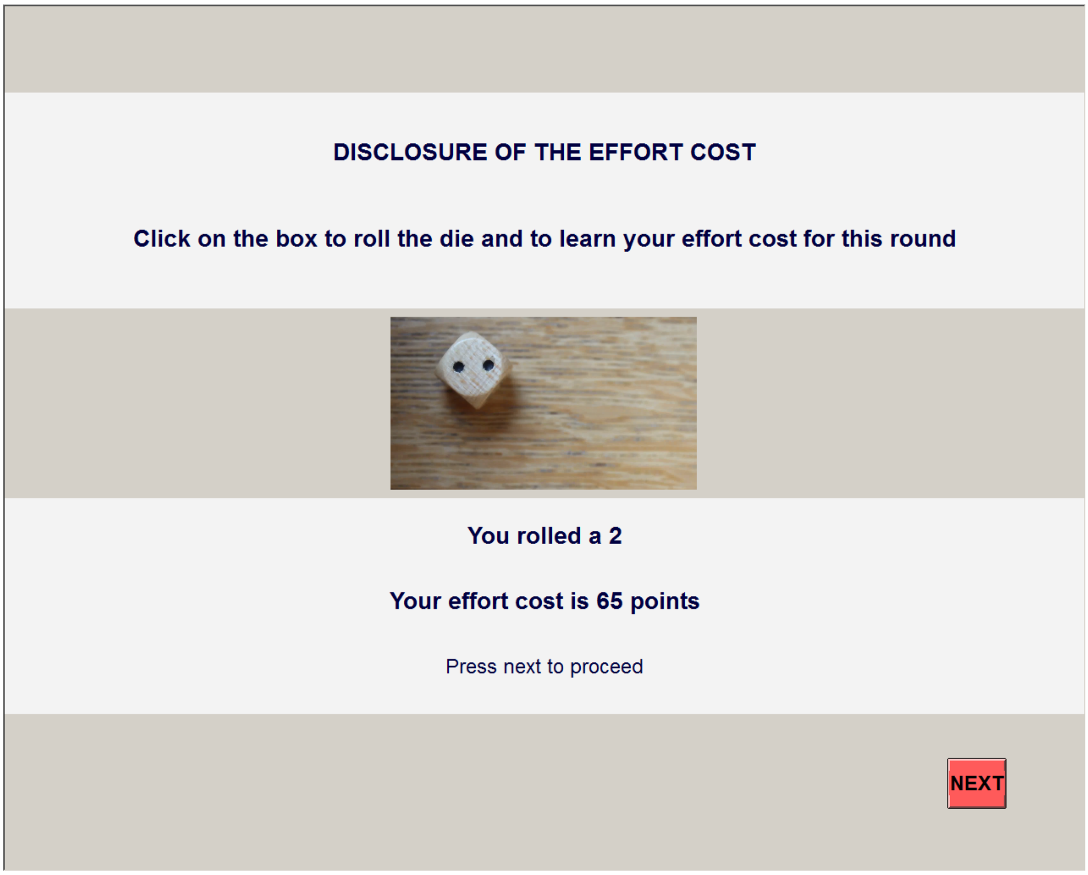
Sequence of Events

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.

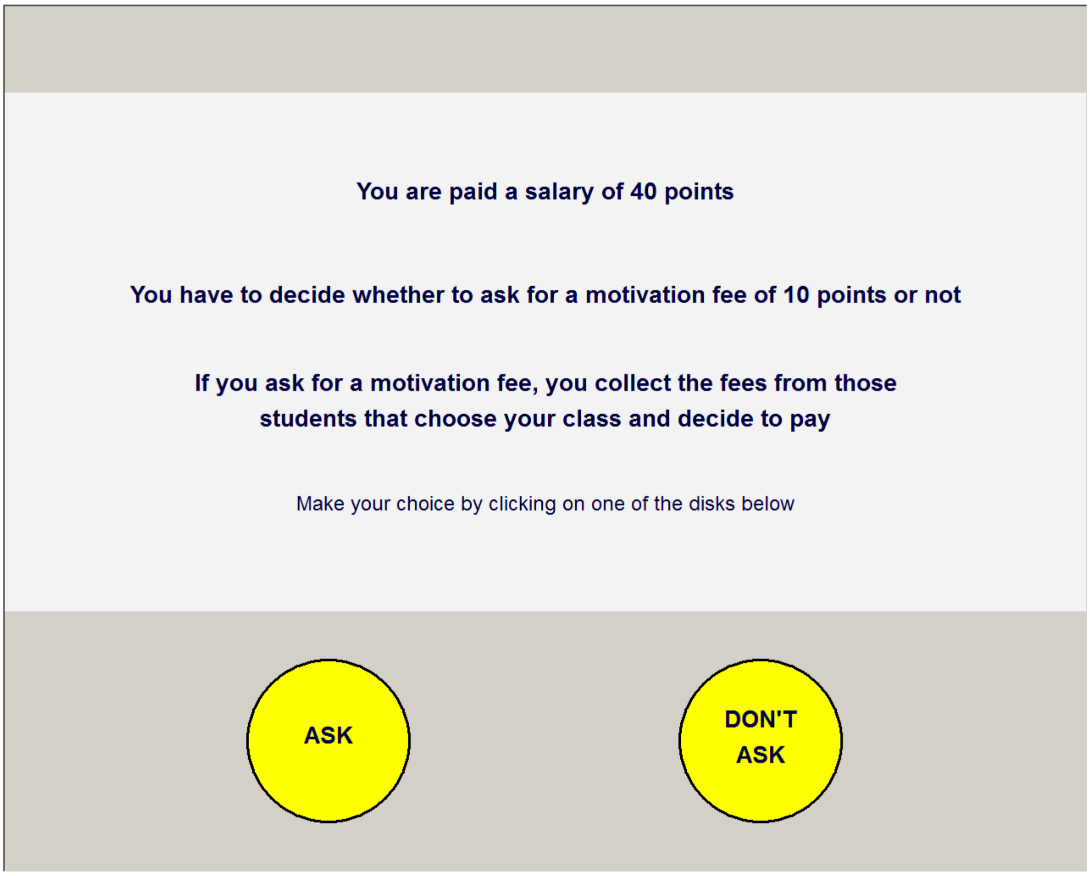


Decisions

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



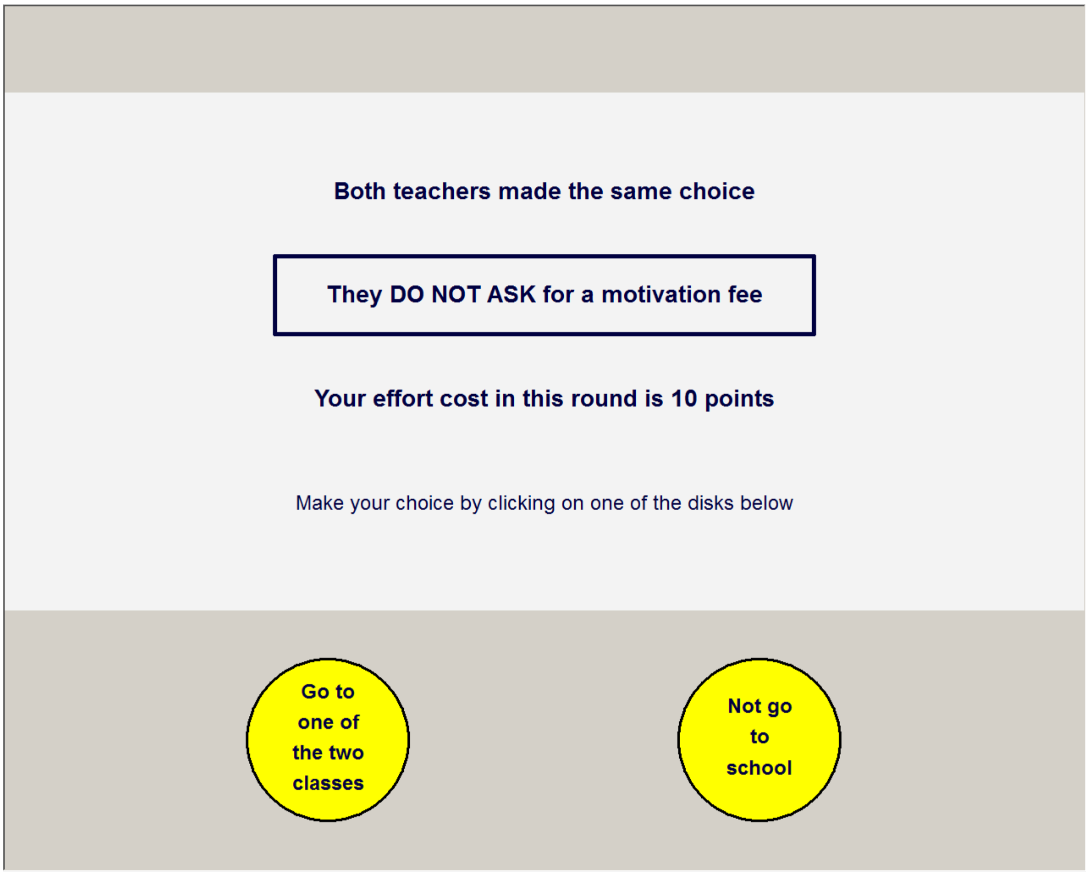
Students' decision

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

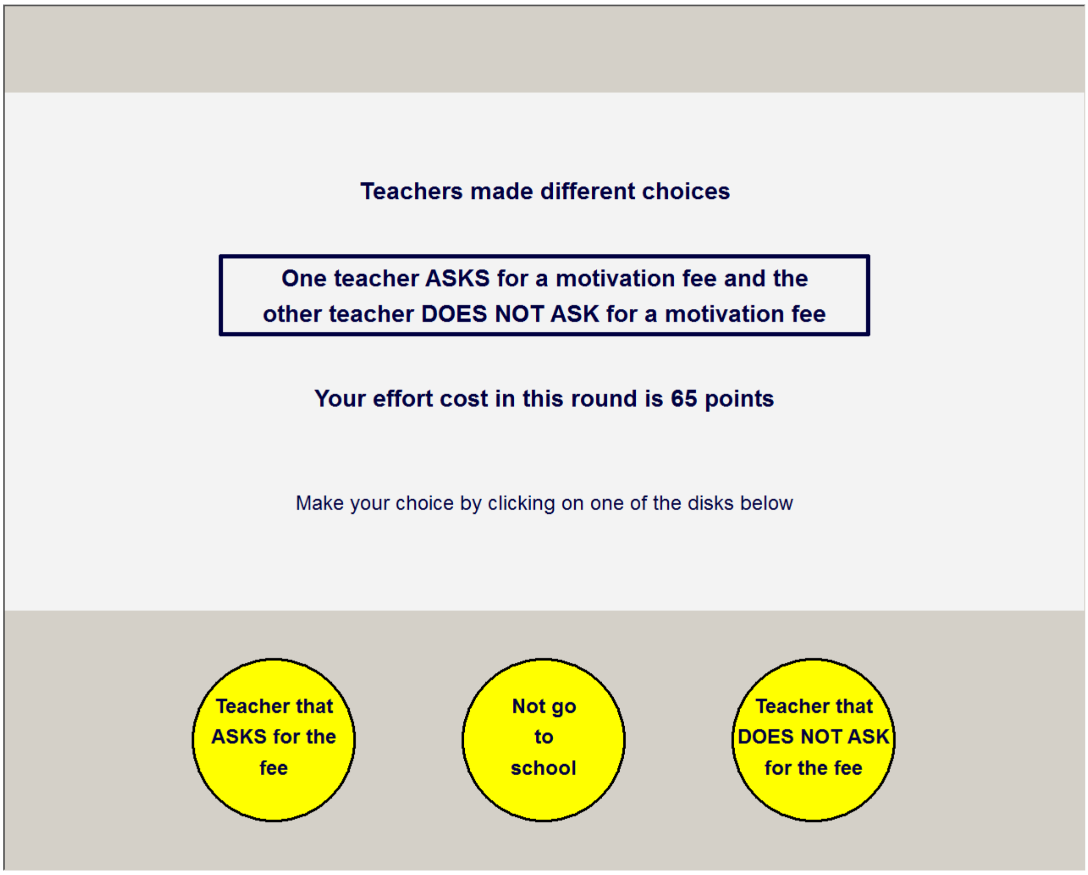
1. not go to school OR
2. be assigned to one of the two teachers and obtain the diploma by exerting effort. The computer will randomly assign half of the students choosing this option to one teacher and the other half to the other teacher.

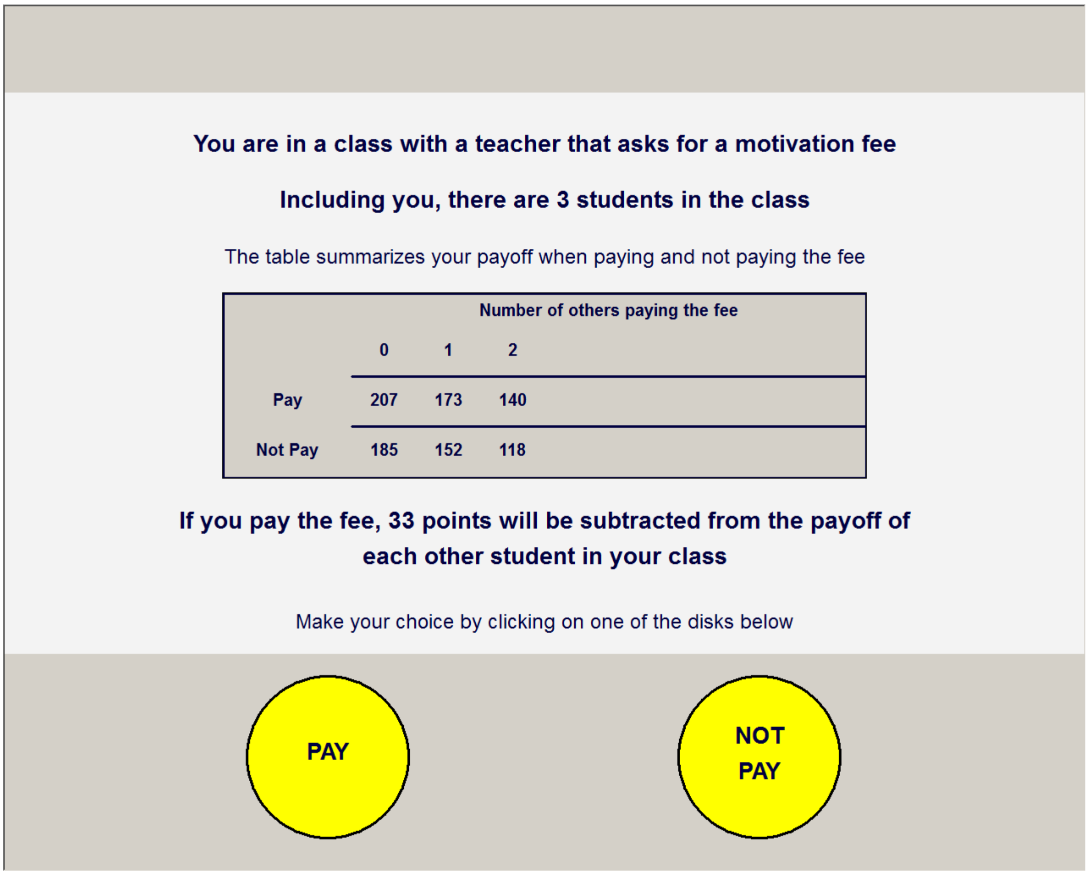


2. One teacher ASKS for a motivation fee and the other teacher DOES NOT ASK for a motivation fee

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



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.

D Inequity averse teachers with the experimental parameters. [ONLINE APPENDIX]

In this section we derive prediction for the model used in the experiment assuming that teachers are inequity averse. To do that we assume that the teachers have a Fehr and Schmidt (1999) utility function, i.e.,

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

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.

D.1 Inequity averse teachers under fixed-wage (*F=40* and *F=240*)

We first derive the utility function of the teachers and consider the following 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 medium students pay the bribe.

Let n_{Gi} be the number of good students in the class of teacher T_i, the payoff of all the students in that class is $\pi_s(n_{Gi}) = 250 - 10 - \frac{100 \cdot (4 - n_{Gi})}{4}$ and the payoff of 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, given n_{G1} and n_{G2} the utility of a teacher when both solicit bribes is

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

and the expected utility is

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

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 solicit bribes and the other teacher do not solicit bribes. When one teacher solicits bribes and the other does not, the good and medium students go to the teacher not soliciting bribes and obtain a payoff $\pi_G = 250 - 10$ and $\pi_M = 250 - 65$, respectively. The bad students go to the teacher soliciting bribes and pay the bribe obtaining the payoff $\pi_B = 250 - 10 - 100$.

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 $\pi_1(B) = F + n_B \cdot 10$ and the teacher not soliciting bribes (T_2) obtains a payoff $\pi_2 = F$. The utility of T_1 given n_B and n_G is

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

and the 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

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

for $F = 40$ and

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

for $F = 240$

Similarly, one can define the utility given n_G and n_B for the teacher not soliciting bribes and then compute his/her expected utility $EU_{(0,1)}(\alpha,\beta)$ obtaining the following

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

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 medium students go to school and obtain a payoff of, respectively, $\pi_G = 250 - 10$ and $\pi_M = 250 - 65$ independently 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 $\pi_i = F$ independently 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 given n_B and n_G is

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

and the expected utility is

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

that, after simplification, 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	
		$B=1$	$B=0$
T_1	$B=1$	$\frac{220}{3}-\frac{799755}{209952}\beta-\frac{140767775}{209952}\alpha$	$\frac{160}{3}-\frac{40}{3}\beta-\frac{3200}{3}\alpha$
	$B=0$	$40-\frac{3560}{3}\alpha$	$40-1140\alpha$

Note that T_1 prefers $(1,1)$ over $(0,1)$ when

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

and $(1,0)$ over $(0,0)$ when

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

These inequalities identifies the regions with different equilibria of the game that are reported in panel (a) of Figure D.7.

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

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

Note that T_1 prefers $(1,1)$ over $(0,1)$ when

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

and $(1,0)$ over $(0,0)$ when

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

These inequalities identifies the regions with different equilibria of the game that are reported in panel (b) of Figure D.7.

D.2 Inequity averse teachers under piece-rate

We first derive the utility function of the teachers. Note that the payoffs of the students remain the same as in the fixed-wage cases. The payoffs that can change are the ones for the teachers that now depend on the number of students in the class. We consider the cases:

(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 of T_i is $\pi_i(n_{Gi}) = 40 + 4 \cdot 50 + 10 \cdot (4 - n_{Gi})$ which is equal to the payoff teachers obtain in the fixed-wage scenario with $F = 240$. Therefore the expected utility of a teacher when they both solicit bribes is the same as in fixed-wage with $F = 240$.

(ii) One teacher solicit bribes and the other teacher do not solicit bribes. Compared to the fixed-wage case, the payoffs of the two teachers change. They now depend on the number of bad students. 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 $\pi_1(n_B) = 40 + n_B \cdot 50 + n_B \cdot 10$ and the teacher not soliciting bribes (T_2) obtains a payoff $\pi_2(n_B) = 40 + (8 - n_B) \cdot 50$. The utility of T_1 given n_B and n_G is

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

and the 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 utility given n_G and n_B for the teacher not soliciting bribes and then compute his/her expected utility $EU_{(0,1)}(\alpha, \beta)$ obtaining the following

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

(iii) Both teachers do not solicit bribes. This case is slightly different than the previous cases. Let n_G and n_B be the number of good and bad students and note that when both teachers do not solicit bribes they share the good and medium students. This means that when a teacher has $\left\lceil \frac{8-n_B}{2} \right\rceil$ students in his class the other teacher has $\left\lfloor \frac{8-n_B}{2} \right\rfloor$ students in his 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 given n_B and n_G is

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

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

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

with probability $\frac{1}{2}$. The expected utility is

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

that, after simplification, 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 of the teachers for 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.

		T_2: $B = 1$	T_2: $B = 0$
T_1	$B = 1$	$\frac{820}{3} - \frac{196754975}{209952}\beta - \frac{799775}{209952}\alpha$	$120 - \frac{2151595}{26244}\beta - \frac{22796875}{26244}\alpha$
	$B = 0$	$\frac{1120}{3} - \frac{183564625}{104976}\beta - \frac{206545}{104976}\alpha$	$\frac{620}{3} - \frac{36176125}{139968}\beta - \frac{9115645}{139968}\alpha$

Note that T_1 prefers $(1,1)$ over $(0,1)$ when

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

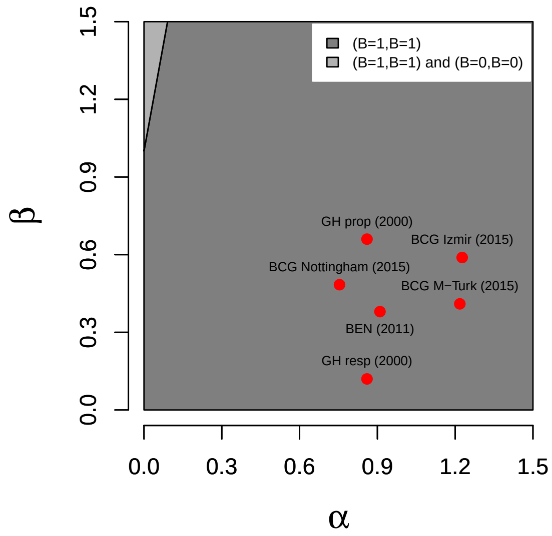
and $(1,0)$ over $(0,0)$ when

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

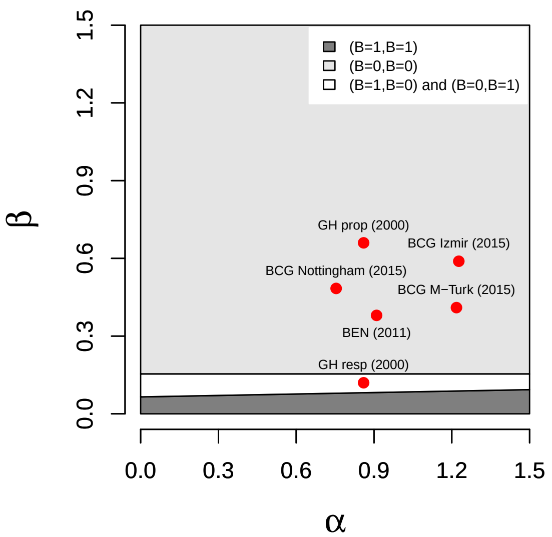
These inequalities identifies the regions with different equilibria of the game that are reported in panel (c) of Figure D.7.

Figure D.7: Equilibria in part 1 and part 2 of the two treatments when teachers are inequity averse with Fehr and 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 and 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.

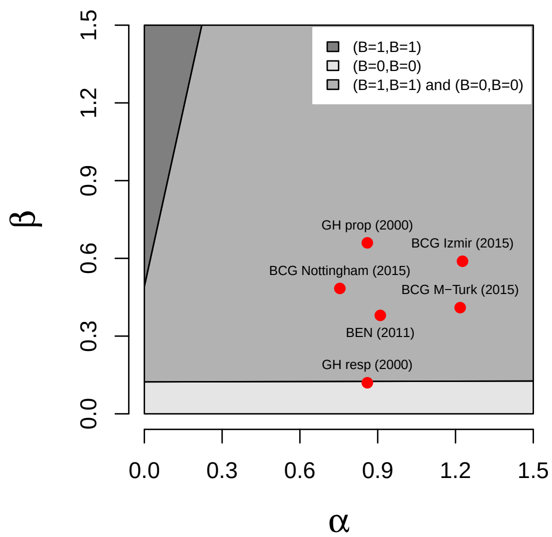
(a) FW and PR - Part 1



(b) FW - Part 2



(c) PR - Part 2



E Additional analysis [ONLINE APPENDIX]

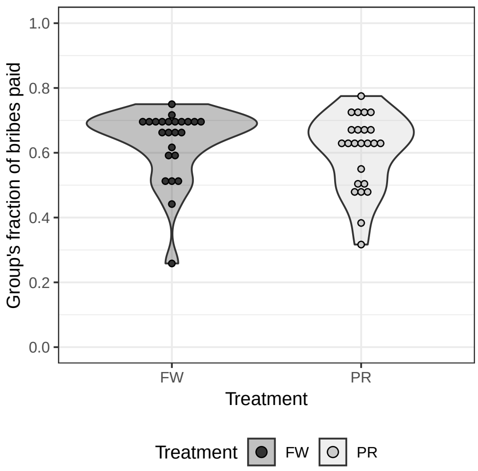
E.1 Overall bribing.

Figure E.1 reports information regarding fraction of bribes paid. Note that bribes are not paid either because the student 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. The fraction of bribes paid in part 1 does not differ in the two treatments. The Wilcoxon rank sum test does not reject the null hypothesis that the median fraction is the same in the two treatments ($p = 0.337$). 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 $p = 0.170$).

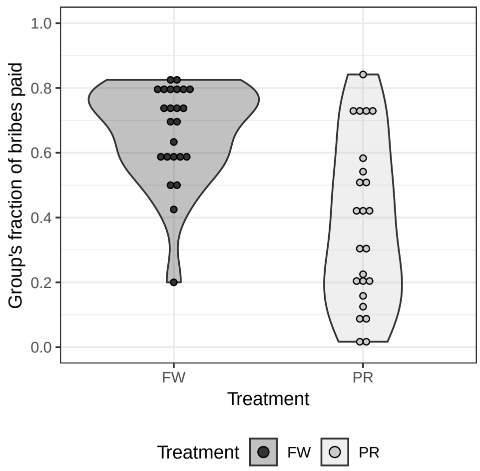
Panel (c) of Figure E.1 differs from panel (c) of Figure 2 because, along with the observed fraction of bribes paid (dots), it reports the theoretical fraction of bribes that should have been paid given the choice of the teachers and the ability of the students (crosses). With fixed-wage the actual frequency of bribes paid is slightly lower than the theoretical fraction.

Figure E.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)

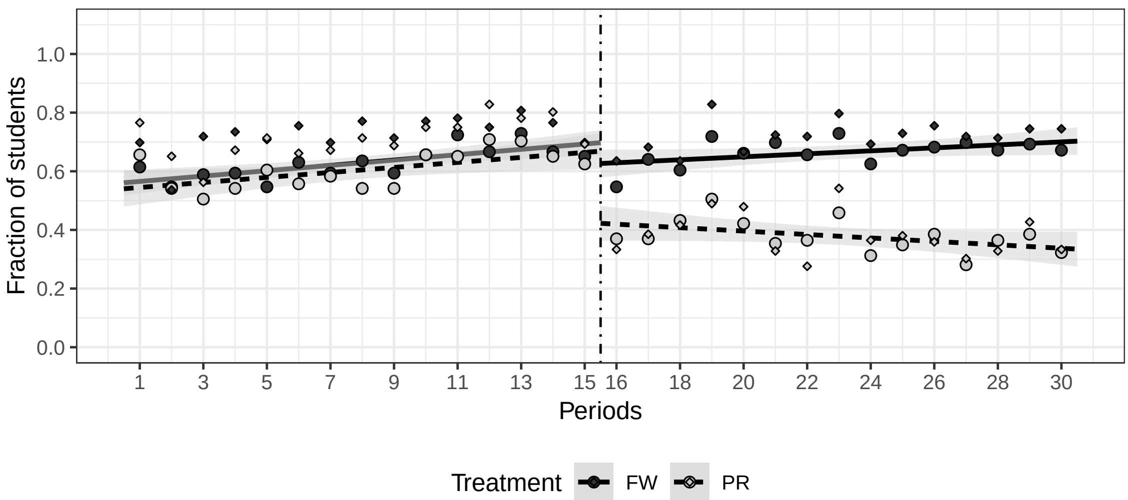
(a) Part 1



(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)

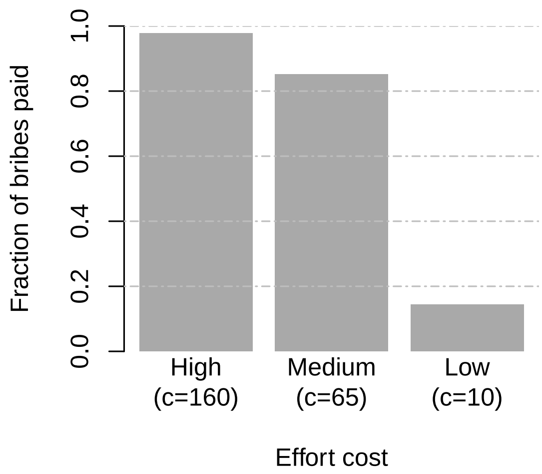


E.2 Students individual behavior.

E.2.1 Choice to pay the bribe or not

Figure E.2 shows the fraction of bribes paid by effort cost. These frequencies are close to the predicted frequencies of 1, 1, and 0, respectively.

Figure E.2: 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.

E.2.2 Choice to go to school

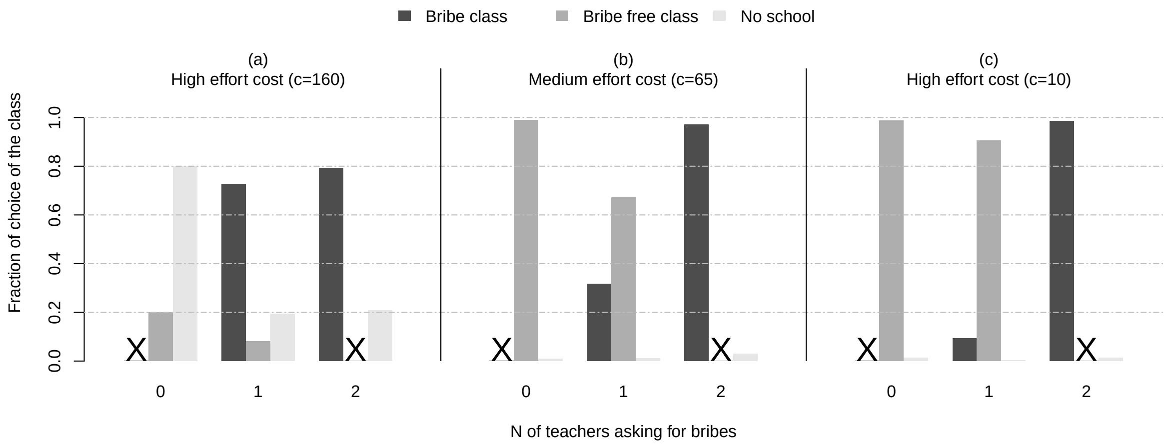
Figure E.3 reports the students' choice of the class by effort cost and by number of teachers soliciting bribes. When the number of teachers soliciting bribes is 0 the choice is between the "Bribe class" class and "No school". When the number of teachers soliciting bribes is 2 the choice is between the "Bribe free class" class and "No school". When the number of teachers soliciting bribes is 1 the choice is among the three options.

Recall that the equilibrium predictions are as follows:

- when the number of teacher soliciting bribes is 0, students with a high effort cost should choose "No school" and the other students should choose the "Bribe Free" class.
- when the number of teacher soliciting bribes is 1, students with a high effort cost should choose the "Bribe" class and the other students should choose the "Bribe Free" class.
- when the number of teacher soliciting bribes is 2, for all the effort costs students should choose the "Bribe" class.

As for the low effort cost, panel (c), the observed fractions of choices are largely in line with the equilibrium predictions. The highest deviation rate occurs when one teacher is soliciting bribes and the other teacher is not soliciting bribes. In this case 9.5% of the choices deviate from the one predicted by the equilibrium.

Figure E.3: Choice of the class by effort cost and options available



For the medium effort cost 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 the "Bribe" class (31.6%). This is difficult to explain even when considering moral concerns. The "Bribe Free" class guarantees a payoff of 185. The "Bribe" class can provide at maximum the same when not paying the bribe and potentially more when paying the bribe and other students do not pay the bribe. Empirically, going to a "Bribe" class and behaving optimally (pay when more than 2 students are in the class and not paying when alone in the class) gave a payoff ≥ 185 only 46 times out of 425 (10.82%).

As for the high effort cost, when the "Bribe" class is available 20% of choices are for "No school". This may be due to a preferences for not paying the bribe (In case of a fully corrupt class of 4 students, if you go to the "Bribe" class and do not pay you get 15 and if you pay you get 140. If you go out you get 115). Moreover, about 20% of the choices are for the "Bribe Free" class when the "No school" option is available (payoff of 90 instead of 115).

E.2.3 Moral vs equilibrium play

Our 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 E.1 summarizes the predicted choices for all possible situations for the two styles of play.

Figure E.4 panel (a) shows the fraction of periods where choices are coherent with equilibrium play and moral play for each individual. Note that the sum of the fractions can be bigger than 1. This is due to the fact that some choice pattern 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 E.4b panel (b) takes into account the fact that predictions of the two styles overlap. In the figures considers only those situations in which equilibrium play and moral play provide

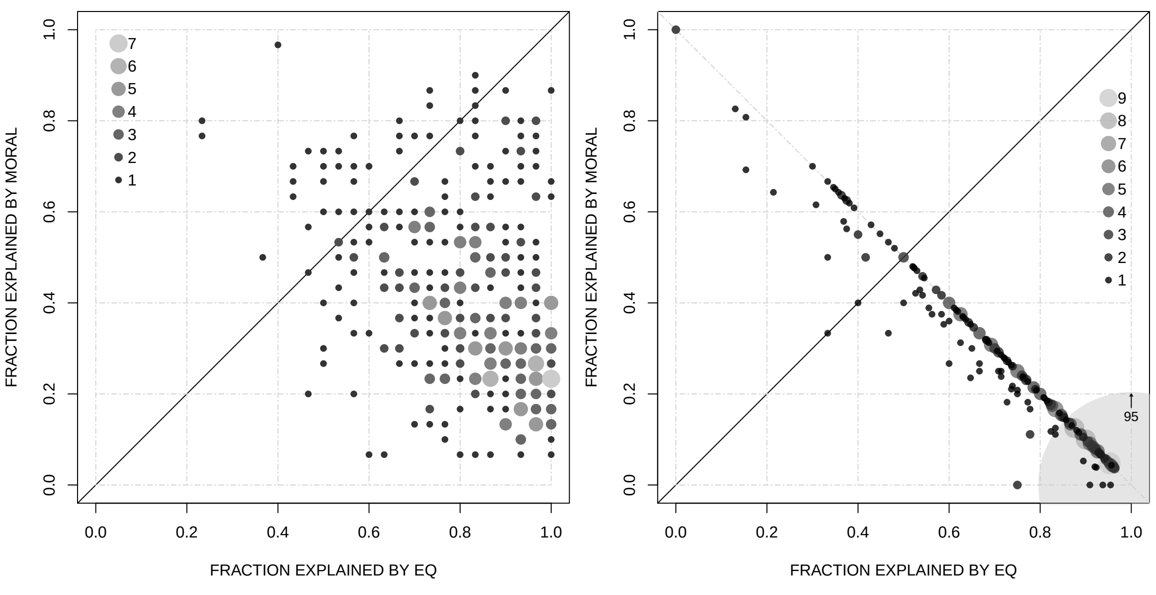
Table E.1: Summary of the choices for the different styles of play (N.S.= no school; B.F.S. = bribe free school (teacher not soliciting); B. S. = bribe school (teacher soliciting))

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 E.4: 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 f (b) only the choices where the predictions differ.

(a) All choices

(b) Moral ≠ equilibrium



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. Compared to the previous picture the sum of the fraction of choices explained by the two styles cannot be greater than one. The figure confirms 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.