A Controlled Field Experiment on Corruption

Olivier Armantier* Amadou Boly†

November 2010

Abstract

This paper reports on a controlled field experiment on corruption designed to address two important issues: the experimenter’s scrutiny and the unobservability of corruption. In the experiment, a grader is offered a bribe along with a demand for a better grade. We find that graders respond more favorably to bigger bribes, while the effect of higher wages is ambiguous: it lowers the bribe’s acceptance, but it fosters reciprocation. Monitoring and punishment can deter corruption, but we cannot reject that it may also crowd-out intrinsic motivations for honesty when intensified. Finally, our results suggest several micro-determinants of corruption including age, ability, religiosity, but not gender.

Keywords: Corruption, Experimental Economics, Field Experiments.
JEL Classifications: C91, C93, D73, I20.

*Federal Reserve Bank of New York, CIRANO, and CIREQ. E-mail: olivier.armantier@ny.frb.org.
†United Nations Industrial Development Organization, Vienna International Center, P.O. Box 300, A-1400, Vienna, Austria. E-mail: a.boly@unido.org.

We are very grateful to Omar Al-Ubaydli, Gary Charness, Jim Engle-Warwick, Gérard Gaudet, Glenn Harrison, Julie Heroux, Abraham Hollander, John List, Claude Montmarquette, Soiliou Daw Namoro, and Clara Sheets. We also thank seminar and conference participants at the Federal Reserve Bank of New York, SUNY Stony Brook, UC Santa Barbara, NYU, the Federal Reserve Bank of Boston, the ESA 2007 meeting in Tucson, the CSAE conference 2008 in Oxford, the IMEBE 2008 in Alicante, the New Techniques in Development Workshop 2008 in Canberra, the CEA 2008 in Vancouver, and the SCSE 2008 in Montebello. We would like to thank the recruiting firm Opty-RH for its help with hiring graders in Burkina Faso. The authors gratefully acknowledge financing from the CRSH. The views expressed here do not necessarily reflect those of the Federal Reserve Bank of New York or the Federal Reserve System. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the United Nations Industrial Development Organization.
1. Introduction

Corruption, defined in economics as an abuse of public office for private gain, is now recognized as one of the most detrimental factors to economic and social development.\(^1\) The analysis of corruption, however, has been a challenge to applied economists. Indeed, because of its illegal and secretive nature, hard data on corruption are virtually inexistent. Recently, laboratory experiments have been conducted in order to overcome the unobservability of corruption. This approach however, raises new challenges. In particular, the essence of corruption may be difficult to capture in the lab. In addition, lab subjects, knowing they are being observed, may alter their behavior. In this paper, we report on a controlled field experiment combining two original features: First, corruption is unambiguously observed; second, subjects act without knowing they participate in an experiment.

Although economists have studied corruption since at least the early 1970’s, the experimental approach to corruption is recent and not yet fully mature.\(^2\) To date, most experiments have been conducted in the lab in order to test bribery deterrents (e.g. penalties, better wages) and identify micro-determinants of corruption (e.g. gender).\(^3\) In the basic bribery experiment a briber chooses an amount to transfer to a bribee. If accepted, the offer is augmented and the bribee decides how much to return to the briber. In most experiments, the transfer may be randomly detected in which case the briber and/or the bribee

\(^1\)In particular, the World Bank states on its website that it “has identified corruption as the single greatest obstacle to economic and social development” (http://www1.worldbank.org/publicsector/anticorrupt/index.cfm).

\(^2\)For general surveys on corruption see Bardhan (1997), Jain (2001), Aidt (2003), or Seldadyo and Haan (2006). For surveys of laboratory experiments on corruption see Dusek, Ortmann and Lizal (2005), as well as Abbink (2006).

is penalized.\textsuperscript{4} Because lab experiments do not involve an explicit abuse of public office for private gain, it may be argued that the essence of corruption has not been captured in the lab. In addition, it is unclear how accurately corrupt behavior may be measured in the lab. In particular, Levitt and List (2007) argue that, because they know they are being observed, lab subjects may be more inclined to make the moral choice when morality and wealth are competing objectives.

We are only aware of two published field experiments directly related to corruption.\textsuperscript{5} The first study, conducted by Olken (2007) in Indonesia, suggests that, unlike grass-root participation, audits can curb-down embezzlement in road construction projects. The second study, conducted by Bertrand et al. (2006) with subjects applying for a driver license in New Delhi, produced mixed evidence to support the “grease-the-wheel” hypothesis (i.e. corruption improves welfare).\textsuperscript{6} While financially incentivized subjects obtained their licences faster, an ex-post driving test indicates that many unsafe drivers got their licenses, while good candidates were denied. Note that in both of these studies, corruption is not measured directly. Instead, it is inferred from observing how the experimental output changes when moving to a treatment in which corruption is considered more likely.\textsuperscript{7}

The controlled field experiment we propose reproduces a corruption scenario in which a grader is offered a bribe with a demand for a better grade. In short, subjects recruited for a

\textsuperscript{4}Several variations of this basic design have been introduced, e.g. some experiments use loaded instructions (Abbink and Hennig-Schmidt 2006), while in others acceptance of the bribe generates a negative externality on a third party (Barr and Serra 2009).

\textsuperscript{5}Some empirical studies have been conducted with data obtained after natural experiments (see e.g. Di Tella and Schargrodsky 2003, or Reinikka and Svensson 2004). We do not consider these to be field experiments as the researchers had no control over the design of the experiment. Recent unpublished field experiments on corruption include Castillo et al. (2009).

\textsuperscript{6}The “grease-the-wheel” hypothesis was proposed by political scientists (e.g. Leff 1964) who suggested that corruption may help overcome systemic governmental inefficiencies and distortions.

\textsuperscript{7}For instance, shorter delays in obtaining a driver license in a treatment in which subjects are given a financial incentive, are attributed to corruption.
part-time job were asked to grade a set of 20 exam papers. The 11th paper came with a money offer and a message stating: “Please, find few mistakes in my exam paper.” The experiment was conducted in Burkina Faso where corruption is widespread, and where, like in our experiment, part-timers are regularly hired to grade exams. Compared to previous corruption experiments, our design combines two distinctive features: i) subjects were informed they participated in an experiment only after grading was completed, and ii) acceptance of the bribe was directly observed in a controlled environment. We conducted different treatments by varying factors susceptible to promote or deter corruption: i) the amount of the bribe, ii) the wage paid to graders, and iii) the level of monitoring and punishment. To the best of our knowledge, no comparable treatments have been compared in a field experiment.

The remainder of the article is organized as follows. The design of the experiment is presented in section 2. The different treatments are explained in section 3. The experimental results are reported in section 4. Finally, section 5 concludes.

2. The Experimental Design

The experiment took place in Ouagadougou (Burkina Faso) during the national exams’ period. Burkina Faso is a developing country in West Africa with over 13 million inhabitants, among which around 1.4 million live in the capital Ouagadougou. With a “Corruption Perception Index” consistently below 5 (3.5 in 2008), this former French colony is considered to have a serious corruption problem. All sectors of the economy seem to be affected by corruption. In particular, the educational sector was ranked in 2003 as the 6th (out of 10)

---

8 National exams must be passed to move from primary to middle school, middle to high school, and high school to college. The exam period typically lasts from June to the end of July.
9 Ironically, the name “Burkina Faso” means the “Land of Incorruptible People.”
most corrupt public sector in the country.\textsuperscript{10}

In Burkina Faso, public and private schools regularly hire part-timers to grade exams.\textsuperscript{11} Following a similar recruiting practice, we used a local recruiting firm (\textit{Opty-RH}) to place flyers around Ouagadougou proposing a part-time job consisting in grading exam papers. The offer stated that the job would pay up to 5,000 \textit{FCFA},\textsuperscript{12} and that it would consist of two sessions: a grading session lasting half a day, followed within a period of ten days by a debriefing session during which graders would be paid.\textsuperscript{13} In addition to a form of identification, the graders had to possess a university diploma or a proof of enrollment at a university. People interested in the part-time job were asked to come register in person at the recruiting firm location. After validating their credentials, graders were given the day, the time, and the location of their two sessions. At no point were the subjects informed they were about to participate in an experiment.

The task given to graders consisted in spell-checking “dictées.” A dictée is a classic exam in the Francophone schooling system, whereby candidates type a text as it is continuously dictated to them.\textsuperscript{14} The text, based on a newspaper article in French, has 290 words and

\begin{flushright}
\textsuperscript{10}For additional information on the extent of corruption in Burkina Faso see the “Etat de la Corruption au Burkina Faso, Rapport 2005” published by the Réseau National de Lutte Anti-Corruption.

\textsuperscript{11}In fact, 90 out of the 179 subjects we interviewed reported having previously taken part in similar part-time grading jobs.

\textsuperscript{12}The Franc CFA is the currency used in Burkina Faso. The conversion rate was roughly $\textdollar{}1 for 477 \textit{FCFA} (or \texteuro{}1 for 656 \textit{FCFA}) at the time the experiment was conducted.

\textsuperscript{13}One of our concerns was to prevent contamination within the subject pool. In particular, we did not want the word to spread that the grading task was not legitimate or that bribes were present in the exam papers. We took several steps in order to prevent these problems: i) we conducted the experiment in a large city; ii) we tried to recruit a geographically diverse subject pool; iii) we conducted all the sessions within a ten day period; iv) we alternated sessions with and without a bribe; and v) subjects were only informed they participated in an experiment during the debriefing sessions, after all the grading sessions had been carried out. Informal conversations suggest that, until it was revealed to them in the debriefing session, the subjects believed they were hired for an actual grading job.

\textsuperscript{14}The aim is to evaluate one’s spelling ability and knowledge of the French grammar (many words in French have the same pronunciation but different spelling). Such a test is administered several times a year to students between the ages of 8 and 14. In addition, it is one of the requirements to obtain a secretary’s diploma, and it is part of the entry-exam to several civil servant positions. Although traditionally conducted
fits on two pages.\textsuperscript{15} To type some of the exam papers, we recruited 23 subjects (called “candidates” hereafter) in Montréal whose payoffs depended i) on the number of mistakes subsequently identified, and ii) on their decision to offer graders a bribe. Since the role of the candidates is essentially passive, the description of the typing sessions and the candidates have been moved to the Appendix.

To control the distribution of mistakes, we selected only 7 out of the candidates’ 23 papers.\textsuperscript{16} Out of these 7 papers, we chose a “bribe paper” with 20 mistakes. Note that this bribe paper had been typed by a candidate who chose to bribe the graders. To complete the set of 20 exam papers, we made up 13 papers with various numbers of mistakes.

As the papers would be graded in a specific order, we ordered the set of 20 exam papers in a precise way. First, we decided to place the bribe paper in 11\textsuperscript{th} position. Second, we ensured that the first and last set of 10 papers each has a symmetric and roughly identical distribution of mistakes. In particular, they have the same average (15.5), the same median (15.5), and roughly the same standard deviation (6.8 versus 6.7).\textsuperscript{17} Third, we decided on a passing grade of 15 mistakes, meaning that if all mistakes were detected and reported, then half of the papers (including the bribe paper) would fail.\textsuperscript{18}

The exam papers were only identified by a 10-character code combining digits and letters. The first two digits, going from 01 to 20, identified the order in which the graders were asked with pen and paper, a dictée may also be typed. In particular, secretary’s exams are still often typed on a typewriter and spell-checked by graders as in our experiment.

\textsuperscript{15} A copy of the materials used for the experiment may be found on the website of one of the authors, https://sites.google.com/site/olivierarmantier/.

\textsuperscript{16} After we eliminated the papers with too many skipped words and too many mistakes, the selection was made on the ground of convenience to generate an appropriate distribution of mistakes.

\textsuperscript{17} Observe that by focusing exclusively on the graders’ behavior we were able to maintain considerable control over the experiment. In particular, we were able to fix the amount of the bribe, the distribution of mistakes in the 20 exam papers, and the number of mistakes in the bribe paper.

\textsuperscript{18} Such a failure rate is common in most exams and admission tests in Francophone countries.
to grade the papers. Finally, we added a front page so as to look like a legitimate exam. This front page included in particular the identification code, as well as the instructions given to the candidates.

To maintain the appearance of an actual grading task, the sessions were conducted in a high school we rented in the center of Ouagadougou. Upon arrival, the subjects were gathered in a large room. Instructions on how to grade the exam papers were read aloud, followed by questions. Each grader was then randomly assigned to a private room where he found an envelope containing the 20 exam papers properly ordered, a report sheet, a red pen, and an answer book (i.e. a copy of the text without mistake). No information was given about the nature of the exam or the candidates. The graders were explicitly instructed to grade the papers in the proper order. After spell-checking a paper, the graders had to report the number of mistakes both on the front page of the paper and on the report sheet. Graders were told that a candidate would fail the exam when more than 15 mistakes are reported. In such cases, we asked the graders to check the “Fail” column on the report sheet next to the number of mistakes. The graders were also instructed not to leave their room under any circumstance until they were done grading the 20 papers. We told them that we would stop by their room every 15 minutes precisely to answer any potential question. Grading therefore took place behind closed doors, and the graders knew they would be undisturbed except at regular 15 minute intervals. Once their task completed, we gave the graders an “IOU”, and reminded them to come back for the debriefing and payment session.

To introduce the bribe, we handwrote “Please, find few mistakes in my exam paper”

19 Note that every feature of the experiment (from the recruiting, to the design of the task, to the way the grading sessions were conducted) were chosen so as to appear like a regular part-time grading job in Burkina Faso. None of the subjects, even those with prior experience with such part-time grading jobs, questioned the credibility of the experiment.
on an easily removable “post-it”, and we taped it with a banknote on the second page of paper.\textsuperscript{20} We made sure that the message and the money were i) attached securely, ii) not visible unless the exam paper was opened to the second page, and iii) discovered before the grader started spell-checking the paper.\textsuperscript{21} When the graders reported the bribe attempt during one of our visits, we asked them to write in bold “fraud attempt” on the paper. We took the banknote and the message, and we instructed the graders to spell-check the bribe paper just like any other paper. Note that the graders knew the implications of reporting the bribe. Indeed, the instructions given to the graders at the beginning of the session mentioned that any attempt at fraud by a candidate would be penalized by failure of the exam. This information was also available in bold on the front page of each paper.\textsuperscript{22}

In the debriefing sessions, graders were first informed that they took part in an experiment. We described the nature of the research and we explained why they were not informed earlier. We also assured the subjects that the decisions they took were inconsequential and that the data collected would be fully anonymized. After answering questions and addressing concerns, the subjects were offered the possibility to sign an ex-post consent form. They were told they could refuse in which case they would still be paid, and their data would be destroyed. We made sure the subjects had the time and the privacy to make an informed

\textsuperscript{20}This approach prevents face-to-face communication and informal bargaining which could have been influenced by the briber’s personal characteristics (e.g. gender, ethnicity). Note also that bribing a grader is not uncommon in Burkina Faso. For instance, a Burkinabe’s newspaper (“Le Pays”) reported on March 7, 2006 that two students were caught in a bribery attempt comparable to the one in our experiment. Incidentally, the first author was once offered (and obviously declined!) $100 as a “Christmas gift” by a first year graduate student in his class, before the grades had been turned in.

\textsuperscript{21}Recall that an exam paper consists of three pages: a front page, plus two pages of text. The bribe and the message were therefore attached to the first page of text. Pictures of the bribe paper, as well as pictures of the high-school where the experiment took place are available on one of the author’s webpage at https://sites.google.com/site/olivierarmantier/.

\textsuperscript{22}Graders asking why they still had to grade the bribe paper despite the candidate’s certain failure, were told that the official policy was to grade every paper.
decision without being pressured. All subjects, in all of the treatments accepted to sign the ex-post consent form.\textsuperscript{23} Finally, the subjects filled a short questionnaire, after which they were paid in cash.

To conclude this section, we must acknowledge that our experiment raises at least two moral and ethical concerns. First, the subjects were not asked for voluntary consent before they took part in the experiment. Although this approach remains controversial, subjects in field experiments are often not informed of their involvement in an experiment, as it is the only way to totally remove scrutiny’s effects.\textsuperscript{24} Second, although a subject in Montréal typed the bribe paper and accepted to send the graders a bribe, some elements of the corruption attempt were under the experimenter’s control (e.g. the amount of the bribe). As a result, our experiment may be construed as involving deception. Although we do not condone such an approach, we note that it is virtually inevitable when conducting a controlled field experiment about an illegal activity. In particular, field experiments on discrimination in economics have long relied on such a deceptive approach.\textsuperscript{25} In contrast with most field experiments, however, we took several precautions to minimize adverse effects on subjects. First, we created a private and closed environment so that our subjects’ decisions entailed no risk or consequence. Second, subjects were carefully debriefed and offered the possibility to withdraw from the study. Third, subjects were given the opportunity to contact (at no cost to them) the researchers and an ombudsman at any point after the experiment to express

\textsuperscript{23}This result may reflect both the fact that petty corruption is common in Burkina Faso, and the interest subjects expressed about our study.

\textsuperscript{24}See e.g. Levitt and List (2007) for citations of such field experiments. See also Levitt and List (2009) for a discussion of ethical issues raised by the absence of voluntary consent in field experiments.

\textsuperscript{25}See e.g. Yinger (1986), Ayres and Siegelman (1995), Neumark et al. (1996), Bertrand and Mullainathan (2004), as well as the numerous references in Riach and Rich (2002). See also Riach and Rich (2004) for a discussion of the ethical issues raised by these experiments.
concerns or talk about their experience. None elected to do so, and we did not find any evidence that subjects were seriously harmed or inconvenienced by the experiment.

3. The Experimental Treatments

Our objective is to better understand factors that may promote or deter corruption. To do so, we conducted six different treatments. In the control treatment, subjects were paid a fixed amount for their grading regardless of their performance. In addition, graders in the control treatment were offered a bribe. The fixed amount, called a wage hereafter, was 5,000FCFA, and the bribe was set at 1,000FCFA. Although we also analyze how a subject’s grading performance is affected after he accepts or rejects the bribe, we will primarily compare the different treatments with respect to two criteria: the frequency with which the bribe is accepted, and the frequency with which graders fail the briber. If graders respond only to financial incentives, then most of them should accept the bribe in the control treatment since they face no explicit risk of being caught. Moreover, graders who accept the bribe should not be expected to fail the briber less often, since they have no direct incentive to reciprocate.

The “High Bribe” Treatment:

The “High Bribe” treatment attempts to address a basic question: Is corruption price elastic? In other words, do bribees respond to higher bribes? Or, is honesty a moral trait

---

26 Note that the wage and bribe amounts were selected so as to be credible. In particular, 5,000FCFA roughly corresponds to the amount our subjects would receive for a similar part-time job of grading.

27 From a legal perspective, the first criterion is the most important. Indeed, virtually all legal systems around the world consider that bribery starts with the acceptance of a bribe. Reciprocation is rarely a necessary condition.
that remains insensitive to price variations? Standard economic theory assumes the former, and several lab experiments support this hypothesis (see e.g. Abbink et al. 2002, Jacquemet 2005, Barr and Serra 2009). To the best of our knowledge, however, there is little field evidence on the price elasticity of corruption. The design of the “High Bribe” treatment is identical to the control treatment except that the amount of the bribe was doubled (i.e. 2,000 FCFA). Our basic hypothesis is that graders should be more inclined to accept the more generous offer. The financial incentives to reciprocate however, remain nonexistent. As a result, graders should fail the briber at the same rate as in the control treatment.

The “High Wage” Treatment:

The objective of the “High Wage” treatment is to test one of the most commonly advocated anti-corruption policies. Indeed, practitioners and international institutions often recommend paying civil servants up to, or even above their private sector alternative as a means to deter corruption. The “High Wage” treatment is identical to the control treatment except that upon arrival at the experiment, the subjects were informed they would be paid 7,000 FCFA instead of the 5,000 FCFA stated on the flyers. We justified this 40% increase above standard market wages by explaining we wanted to provide graders an extra motivation to perform their task as well as possible.

From a theoretical perspective, the fair wage-effort hypothesis (Akerlof 1982) predicts that our graders should respond to the unexpected increase in wage by demonstrating greater

---

28 Singapore and Hong Kong are often presented as successful examples of such a policy. Indeed, these countries are typically ranked among the least corrupt, and they are known to pay high salaries to their civil servants. In particular, the prime minister of Singapore is paid several times more than the U.S. president. See also the United Nations’ “Manual on Anti Corruption Policy” (2006).

29 Consistent with the fair wage-effort hypothesis, we used a surprised bonus to make our subjects aware that they were paid more than normal. We cannot exclude, however, that the justification given for the bonus created a frame that may in part explain our results.
reciprocity toward their employer. In other words, compared to the control treatment, we should expect an increase in both the bribe rejection and the failure rate for the bribe paper. The experimental literature on gift exchange however, does not unambiguously support this prediction. Likewise, the few existing empirical and lab studies produced mixed results as to the impact of higher wages on corruption.

The “Monitoring and Punishment” Treatments:

Two treatments were conducted in an attempt to test the efficacy of the most commonly implemented anti-corruption policy: monitoring and punishment. For obvious practical and ethical reasons, we decided against confronting graders who accept the bribe. Instead, we introduced a mechanism aimed at monitoring grading accuracy. This indirect approach therefore makes it possible to detect and punish corrupt graders when they favor the briber. In the “Low Monitoring” treatment, we told each grader we would randomly pick and regrade 1 of the 20 papers he spell-checked to verify whether they performed their task properly. The penalty depended on the difference between the number of mistakes reported by the grader and the actual number of mistakes in the paper. The penalty imposed was either (a) 1,000, (b) 2,000 or (c) 4,000 FCFA when the difference was respectively between (a) 3 and 5, (b) 6 and 9, or (c) 10 and more mistakes. We also conducted a “High Monitoring” treatment in

---

30While several experiments support the fair wage-effort hypothesis (e.g. Fehr, Gächter and Kirchsteiger 1997), other lab and field studies have questioned the robustness of the results (e.g. Gneezy and List 2006).

31In the empirical literature, van Rijckeghem and Weder (2001), as well as Alt and Lassen (2003) find a negative correlation between wages and corruption, while Rauch and Evans (2000), and Swamy et al. (2001) find no support for this common hypothesis. In the experimental literature, Barr, Lindelow, and Serneels (2004) find a significant negative correlation between wages and corruption, while no such treatment effect has been detected in Abbink (2002) or Schulze and Frank (2003).

32Extreme cases of such policy include China where for the first time a high ranking official (the previous head of the Food and Drug Administration) was executed on July 10, 2007 after admitting to taking bribes. Likewise, Hong Kong created a legal precedent by implementing a “guilty until proven innocent” approach toward individuals accused of corruption.
which we increased both the probability of detection (from 1 to 5 papers randomly regraded) and the three possible penalties (to respectively (a) 2,000, (b) 3,000 and (c) 4,500 FCFA). Only the worst of the 5 papers regraded was considered for the monetary penalty in the “High Monitoring” treatment. As a result, subjects were guaranteed a minimum payment of respectively 1,000 FCFA and 500 FCFA in the low and high monitoring treatments. Except for the risk of being penalized, the two monitoring treatments are otherwise identical to the control treatment.

In our experiment, because of the type of monitoring introduced, we should not expect any change in the acceptance rate of the bribe compared to the control treatment. With the intensification of monitoring and punishment, graders who accept the bribe should be less likely to report fewer mistakes for the brieber, because of the increased risk of being penalized. The experimental evidence on the effectiveness of monitoring, however, is mixed. While some experiments suggest that monitoring may be effective (e.g. Falk and Kosfeld 2006), its undermining effect can sometimes outweigh its disciplining effect (Gneezy and Rustichini 2000, Schulze and Frank 2003). As a result, the “High Monitoring” treatment may not necessarily be more effective than the “Low Monitoring” treatment in curbing-down corrupt behavior.

The “No Bribe” Treatment:

Finally, to test whether the presence of the bribe affects grading, we conducted a “No Bribe” treatment in which no bribe was offered. Although secondary, this treatment will give us the opportunity to better assess whether graders who accept the bribe reciprocate by failing the brieber less often than they normally would. Likewise, we will be able to test
whether subjects who reject the bribe are affected by the corruption attempt when grading the briber’s paper.

As indicated in Table 1, a total of 247 subjects participated in the experiment, with a minimum of 37 subjects per treatment. The subject pool is mostly composed of men (86%) and ranges in age between 20 and 33 (with a median at 25 years old). The subjects took on average two and a half hours to grade the 20 exam papers.

4. Experimental Results

The presentation of the results is divided in three parts: First, we analyze the graders’ decision to accept the bribe. Then, we investigate whether graders who accept the bribe reciprocate by reporting fewer mistakes for the briber’s paper. Finally, we explore whether the decision to accept or reject the bribe affects the way a grader subsequently performs her grading task for the remaining nine exam papers.

4.1. The Decision to Accept the Bribe

We start with a brief presentation of descriptive statistics. Then, we conduct an econometric analysis to test for treatment effects while controlling for possible micro-determinants of corruption.\textsuperscript{33} Table 1 gives the frequency of the bribe acceptance in the different treatments. In the control treatment, 49\% of the subjects accepted the bribe. In other words, half of the graders essentially refused “free money” despite the absence of explicit risks and negative externalities. This high rejection rate contrasts with our basic hypothesis and

\textsuperscript{33}Due to small sample sizes, non-parametric tests have low power and some of the treatment effects only emerge when we control for individual characteristics.
with comparable lab experiments.\textsuperscript{34} We conjecture three potential reasons to explain this difference. First, unlike previous lab experiments on corruption, our subjects thought that the grading task and the bribe were real. Second, since subjects are or used to be university students, they can personally relate to the grading task, and therefore, they may be less tolerant toward this specific corruption attempt. Third, controlling the subjects’ beliefs is more difficult in the field than in the lab. In particular, it is highly unlikely that our graders believed they faced absolutely no risk of being caught. Rejection of the bribe in our experiment is therefore not attributable solely to intrinsic motivations. Table 1 also indicates that doubling the bribe increases its probability of acceptance. In contrast, the bribe is rejected more often when the wage is higher and when the graders’ work is monitored.

To test these treatment effects, we estimated a probit model for the grader’s decision to accept the bribe (Model 1 in Table 2). In addition to treatment dummies, the model accounts for several individual specific covariates. In particular, we control for the grader’s age, gender and religiosity.\textsuperscript{35} To capture a grader’s ability at the grading task, we include two variables measuring the grader’s precision and improvement over the first ten exam papers.\textsuperscript{36} Finally, the econometric model accounts for the time the grader took to complete

\textsuperscript{34}For instance, only 9.4% of the subjects in Frank and Schulze (2000) acted honestly even though corruption had a negative externality on an actual public entity (i.e. a film club). Likewise, the rejection rate was only 13.1% in Cameron et al. (2009) in an environment with negative externalities and possible punishment.

\textsuperscript{35}The measure of religiosity was obtained from the post-experiment survey in which we asked subjects how often they go to a church, a mosque or any other place of worship. This variable has 5 categories, ranging from 0 (never) to 4 (every day). Although we did not ask for religious denominations, note that approximately 60% of Burkinabes practice Islam, while another 20% are Christians.

\textsuperscript{36}Formally, $\text{Precision}_i = \frac{1}{10} \sum_{t=1}^{10} Y_{i,t}$, where $Y_{i,t} = - \left| \frac{R_{i,t} - M_t}{M_t} \right|$, $R_{i,t}$ is the number of mistakes reported by subject $i$ for paper $t$, and $M_t$ is the true number of mistakes in paper $t$. Note that $\text{Precision}_i \leq 0$, and grader $i$ is considered more precise when $\text{Precision}_i$ increases toward 0. To obtain the improvement measure, we estimate for each subject $i$ the regression model $Y_{i,t} = \alpha_{0,i} + \alpha_{1,i} \cdot t$ (for $t = 1, ..., 10$), and we set the variable $\text{Improvement}_i$ equal to the estimated slope $\hat{\alpha}_{1,i}$. Grader $i$ is then considered to have improved at the grading task when $\text{Improvement}_i$ is positive and large. Finally, note that these measures of ability are valid instruments since the graders were unaware of the presence of corruption until they reached paper 11.
the grading task.

The results reported in Table 2 indicate that, consistent with our hypothesis, doubling the bribe increases significantly (by 25%) its probability of acceptance. Although the magnitude is not directly comparable, the direction of the effect is consistent with previous lab experiments (Abbink et al. 2002, Jacquemet 2005). Table 2 also indicates that the rate of acceptance of the bribe drops by 23% after a 40% surprised increase in salary.37 The direction of this effect is consistent with the gift-exchange hypothesis, as well as several lab and empirical analyses. Finally, note that the parameters associated with the monitoring treatments are both negative, thereby indicating more bribe rejections than in the control treatment. The effect however, is only significant in the “Low Monitoring” treatment where the rejection rate of the bribe increases by 27%. It therefore appears that monitoring and punishment can be an effective anti-corruption measure. However, the absence of significant treatment effect when monitoring is intensified is also consistent with the hypothesis that monitoring and punishment may crowd-out intrinsic motivations for honesty.38 The relative effectiveness of monitoring at curbing-down the bribe acceptance is somewhat surprising since the treatments were not designed to catch the graders who accept the bribe. Instead, they were designed to catch corrupt graders who reciprocate by giving a better grade to the briber. We conjecture that the implementation of the monitoring and punishment policy may have reinforced the graders’ beliefs about the probability of being caught accepting the bribe, or about the value of the grading task to their employer.

In terms of micro-determinants of corruption, we find that an older, more religious,

37 This result however is not quite significant at the 5% level. The \( P \)-value is 5.83%.
38 The absence of significant treatment effect in the “High Monitoring” treatment may also simply reflect the lack of statistical power due to the small sample size.
more precise or more deliberate grader is significantly less likely to accept the bribe. The magnitude of the religiosity effect is particularly remarkable. Indeed, our results suggest that a subject going to a place of worship everyday is 58% less likely to accept the bribe than a subject that never attends. Although the robustness of these results remain to be confirmed, it is worth noting that, to the best of our knowledge, this is the first study to identify religiosity as a micro-determinant of corruption. Table 2 also indicates that gender does not influence significantly the decision to accept the bribe. This result is somewhat surprising as previous lab experiments suggest that women are less corruptible (Frank and Schulze 2000, Rivas 2008). It also contrasts with a commonly held belief among practitioners that women are less susceptible than men to accept bribes.\footnote{For instance, the police department in Mexico City decided in 1999 to dispatch women traffic officers at sensitive intersections because they were deemed less corruptible than their male counterparts. See e.g. the August 15, 1999 New York Times article available online at \url{http://query.nytimes.com/gst/fullpage.html?res=940CE7DA1239F936A2575BC0A96F958260&n=Top/Reference/\%20Topics/Subjects/P/Police\%20Brutality\%20and\%20Misconduct}).

4.2. The Decision to Report a Failing Grade.

The descriptive statistics in Table 1 indicate that, except for the monitoring treatments, graders who accept the bribe tend to fail paper 11 less often.\footnote{We concentrate on the most relevant decision made by the grader: whether or not to report a failing grade for paper 11. An analysis based on the actual grade reported for paper 11 yields the same conclusions.} Observe however, that corrupt graders do not systematically report a passing grade. In other words, consistent with our prediction, a non-negligible number of graders act opportunistically by taking the bribe and doing nothing in return.

To better understand the decision to fail the briber we estimated three probit models. The first (model “All” in Table 2) is estimated with the entire sample of graders, including
those in the “No Bribe” treatment. To estimate the other two models (models “Accepters” and “Rejecters” in Table 2), we partitioned the data collected in the treatments with a bribe in two groups of bribe accepters and bribe rejecters. These two additional models allow us to test for treatment effects separately on each group of graders. The estimation results for all three models are reported in Table 2.

To start, observe that the parameters associated with the “Precision” and “Improvement” variables are highly significant in all three models. In other words, as one would expect, failure of the bribe paper is essentially explained by the grader’s ability. The results obtained for the “Accepters” model reveal additional individual specific effects. In particular, we find that within the population of accepters, women and older graders are more likely to reciprocate by awarding a passing grade to the briber. Note also that the parameter associated with “Accept” in model “All” is negative and significant. In other words, we find that despite having no financial incentive to do so, graders who accept the bribe reciprocate by failing the briber 32% less often. In contrast, we find no evidence that the bribe rejecters were affected by the corruption attempt when grading the briber’s paper. In particular, the parameter associated with the “No Bribe” treatment variable in model “All” is not significantly different from zero.

In terms of treatment effects, the results from models “All” and “Accepters” suggest that doubling the bribe increases significantly (by 24% in the “Accepters” model) the probability

---

41 A bivariate probit model for the grader’s decisions to accept the bribe and to fail the briber yields similar results, but it does not allow one to compare the grading quality with the “No Bribe” treatment.

42 The magnitude of these effects is non-negligible. In particular, the results from the “Accepters” model suggest that a woman is 35% less likely to fail the briber than a man (this marginal effect accounts for the influence of the interaction term “Female * Monitoring”).

43 Additional tests confirm that, in terms of failing paper 11, the rejecters in the different treatments cannot be distinguished from the subjects who were not offered a bribe.
that corrupt graders report a passing grade for the briber. Larger bribes therefore promote the bribe acceptance, but they also foster reciprocation.

The estimation outcomes for models “All” and “Accepters” also indicate that accepters are more likely (+25% in the “Accepters” model) to reciprocate in the “High Wage” treatment. In other words, we find the effect of a wage increase on corruption to be ambiguous: it reduces the probability of accepting the bribe, but those who accept tend to reciprocate more often by giving the briber a passing grade. The latter result may reflect self-selection, whereby only the corrupt reciprocators keep accepting the bribe in the “High Wage” treatment.

Finally, the impact of monitoring and punishment seems to differ by gender. Indeed, observe that the parameters of the monitoring treatment variables are only significant in the “All” and “Accepters” models when interacted with “Female.” More specifically, we find that a woman is on average 32% more likely to fail paper 11 when her work is monitored. This result is consistent with Frank and Schulze (2000) and Schulze and Frank (2003), who find that women are more responsive to monitoring and punishment.

4.3. Corruption and Subsequent Performance

We now investigate whether the decision to accept or reject the bribe affects how a subject subsequently grades the remaining nine exam papers. To do so, we estimate two panel data models with the entire sample of graders from all 6 treatments. In the first model, the dependent variable is $R_{i,t}$, the number of mistakes reported by grader $i$ for exam paper $t$ ($t = 12, ..., 20$). In the second model, the dependent variable is the grading precision

\footnote{See Boly (2010) for an analysis of treatment effects and grading quality over the first 10 exam papers.}
\[ Y_{i,t} = -|R_{i,t} - M_t| / M_t \], where \( M_t \) is the actual number of mistakes in exam paper \( t \). In both models, the vector of explanatory variables includes the two variables measuring the initial ability of grader \( i \) over the first ten exam papers (i.e. “Precision” and “Improvement” defined in footnote 36). To test whether grader \( i \) keeps improving as he or she did over the first ten exam papers, we control for the exam number multiplied by the value of the variable “Improvement” for subject \( i \). The explanatory variables also include the other individual specific characteristics (i.e. gender, age, religiosity, grading time), a fixed effect for each exam paper, as well as dummy variables for the decision to accept or reject the bribe in each relevant treatment. The subject of reference is therefore a grader in the “No Bribe” treatment. Finally, to control for possible grader specific random effects, we model the error term as \( U_{i,t} = \varepsilon_i + V_{i,t} \), where \( Var(V_{i,t}) = \sigma^2 \), \( Var(\varepsilon_i) = \sigma_a^2 \) when subject \( i \) is an accepter and \( Var(\varepsilon_i) = \sigma_r^2 \) when subject \( i \) is a rejecter.

The results reported in Table 3 indicate that, unlike the other individual characteristics, “Precision” and “Improvement” are highly significant both in Model 1 and in Model 2. In other words, and according with intuition, most of the variation in the reported number of mistakes over the last nine papers may be explained by the subjects’ initial abilities. The parameter associated with the variable “Improvement * Exam Number” is close to, but significantly greater than zero in both models. This therefore suggests a persistence in improvement, whereby subjects whose initial grading quality improved (respectively, deteriorated) over the first 10 exam papers, keep improving (deteriorating) after the bribe paper.

In terms of treatment effects, we can see in Table 3 that, regardless of the model estimated, the parameters corresponding to the subjects who reject the bribe do not have a consistent
sign and are all insignificant. Therefore, in terms of grading the last nine exam papers, we are unable to distinguish the subjects who rejected the bribe from the subjects who were not offered a bribe. This result reinforces the finding in Section 4.2 showing that the bribery attempt did not affect the grading behavior of subjects who reject the bribe.

In contrast, every parameter associated with the subjects who accepted the bribe is negative in Table 3. Observe, however, that the treatment effects are significant at the 5% level only in Model 2. In other words, it seems that the grading quality deteriorates after a subject accepts the bribe, but we do not find conclusive evidence that the bribe takers report fewer mistakes for the last 9 exam papers. Furthermore, note in Model 2 that the standard error of the individual specific effect is significantly larger for the accepters than for the rejecters (i.e. $\hat{\sigma}_a > \hat{\sigma}_r$). In other words, after controlling for initial ability, we find that corrupt graders become not only less precise, but also more inconsistent when grading the last nine exam papers. Interestingly, this suggests that the cost of corruption in this experiment setting is twofold: Not only is the briber more likely to pass, but subsequent exams are also graded with inferior quality. To explain this result, we conjecture that some accepters may try to appear incompetent in order to conceal their corrupt behavior.

Finally, observe that the accepters in the “High Monitoring” treatment are statistically indistinguishable from the graders in the “No Bribe” treatment in both models. In fact, additional tests reveal that the accepters in the “High Monitoring” treatment do a significantly better grading job over the last nine exam papers than the accepters in other treatments. This result is consistent with payoff maximization, as accepters seem to best respond to the higher level of monitoring in an effort to lower their expected penalty.
5. Conclusion

As argued by several international institutions (e.g. the IMF, the World Bank), corruption is one of the most detrimental factors currently afflicting developing countries. Due in part to its secretive nature, applied economists have had limited success in their effort to understand and combat corruption. This paper reports on a controlled field experiment on corruption designed to address two important methodological issues: the influence of the experimenter’s scrutiny on the subjects’ behavior and the difficulty in collecting hard data on corruption. In short, following a practice used in Burkina Faso, we hired part-timers to grade exam papers. One of the papers came with a bribe offer and a demand for a better grade.

The experimental results suggest several micro-determinants of corruption. In particular, the probability to accept the bribe seems to decrease with the grader’s age, religiosity and ability at the grading task. In contrast with several lab experiments, we find no evidence that gender influences the bribe acceptance. Unlike men however, women seem to respond to monitoring and punishment by failing the briber more often. Our results also indicate that the cost of corruption to the employer (if measured by grading quality) is not limited to the briber. Indeed, we find that graders who accept the bribe i) are likely to reciprocate by reporting fewer mistakes for the briber’s paper, and ii) subsequently become less precise at the grading task. In contrast, it appears that the grading quality of subjects who refuse the bribe is not affected by the corruption attempt.

In terms of treatment effects, we find that doubling the bribe increases its probability of acceptance, but also induces a more reciprocal behavior toward the briber. In contrast, the evidence produced by two of the most commonly advocated anti-corruption measures
are not as conclusive. While a higher wage lowered the probability to accept the bribe, it also reinforced the bribee’s propensity to reciprocate by awarding the briber a passing grade. Moreover, while we find that monitoring and punishment can be an effective anti-corruption policy, we also fail to identify a treatment effect in the “High Monitoring” treatment. This result is consistent with the hypothesis that, when intensified, monitoring and punishment can crowd-out intrinsic motivations for honesty.

Although they illustrate some of the difficulties in combating corruption, our results should be considered encouraging. Indeed, although the “Low Monitoring” treatment was not designed to catch graders who accept the bribe, it proved particularly effective at curbing down both bribe acceptance and reciprocation, especially among women. In addition, some of the results obtained in the “High Wage” treatment may have been driven by the part-time nature of the grading job. Higher wages may provide an even better deterrent against corruption with full-time workers, as they may be unwilling to risk losing their stream of future wages.

**References**


Rauch, J. E., and P. B. Evans (2000): “Bureaucratic Structure and Bureaucratic Perfor-


5.1. Appendix: The Candidates (Montréal, Canada)

Subjects, called “candidates”, were recruited to type a “dictée.” At the beginning of the typing session, each candidate was assigned to an isolated computer. Instructions were then read aloud, followed by questions. We explained what would and what would not constitute a mistake. The subjects were also informed that, at the end of the dictation, they would not be allowed to spell-check or modify their papers in any way. We told the candidates that we would decide whether their paper would be spell-checked by an experimenter, or by various subjects called “graders.” Finally, we explained that a candidate’s payment would depend in part on the average number of mistakes the grader(s) would report. The lower the number of mistakes reported, the higher the payment.

Each candidate was also asked whether he would be willing to send some of the graders a money offer (explicitly referred to as “a bribe”), accompanied by the following message: “Please, find few mistakes in my exam paper.” We explained to the candidates that if they accepted to offer a bribe, then their payoffs may not depend exclusively on the number of mistakes reported. Instead, they may also be affected positively or negatively by each grader’s decision to accept or reject the bribe. Finally, the candidates were informed that even if they agreed to offer a bribe, we would not necessarily send the message and the bribe to the graders.45

To keep as much control as possible over the experiment, we deliberately left the candidates’ instructions partly ambiguous.46 In particular, we did not explain how we would select the papers to be graded by experimental subjects. Likewise, we did not specify the precise

---

45Immediately after reading the instructions, subjects were given the opportunity to leave the laboratory with C$10 without having to type the text. None elected to do so.
46According with standard lab practices, however, we did not explicitly mislead the candidates. We simply withheld some information.
way in which the candidates’ payoffs would be calculated. We also remained ambiguous about the amount of the bribe that would be proposed to the graders. The candidates were told they would receive several payments: C$20 payable immediately after the conclusion of the typing session and additional amounts paid after the completion of the grading sessions. The candidates knew that each of the additional amounts could vary between C$20 and C$60, depending on the average number of mistakes reported, and, when relevant, on the number of graders who accepted and rejected the bribe offer.

We conducted two typing sessions in Montréal at CIRANO’s Bell Laboratory for Experimental Economics. Each session lasted roughly an hour and included respectively 11 and 12 subjects. All 23 subjects agreed to send a bribe to the graders.
## Table 1
### Descriptive Statistics

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Control</th>
<th>High Bribe</th>
<th>High Wage</th>
<th>Low Monitoring</th>
<th>High Monitoring</th>
<th>No Bribe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Subjects</td>
<td>37</td>
<td>45</td>
<td>39</td>
<td>39</td>
<td>44</td>
<td>43</td>
</tr>
<tr>
<td>Share of Graders who Accept the Bribe</td>
<td>0.49</td>
<td>0.69</td>
<td>0.36</td>
<td>0.36</td>
<td>0.41</td>
<td>___</td>
</tr>
<tr>
<td>Share of Graders who Report more than 15 Mistakes for Paper 11</td>
<td>All 0.65</td>
<td>0.42</td>
<td>0.41</td>
<td>0.59</td>
<td>0.68</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>Accepters 0.56</td>
<td>0.35</td>
<td>0.21</td>
<td>0.64</td>
<td>0.67</td>
<td>___</td>
</tr>
<tr>
<td></td>
<td>Rejecters 0.74</td>
<td>0.57</td>
<td>0.52</td>
<td>0.56</td>
<td>0.69</td>
<td>___</td>
</tr>
</tbody>
</table>

## Table 2
### Probit Regressions for the Decisions to Accept the Bribe and to Fail the Briber

<table>
<thead>
<tr>
<th>Variables</th>
<th>Model 1 Decision to Accept the Bribe</th>
<th>Model “All” Decision to Fail the Briber</th>
<th>Model “Accepters” Decision to Fail the Briber by the Bribe Accepters</th>
<th>Model “Rejecters” Decision to Fail the Briber by the Bribe Rejecters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Parameter Estimate</td>
<td>Marginal Effect</td>
<td>Parameter Estimate</td>
<td>Marginal Effect</td>
</tr>
<tr>
<td>Constant</td>
<td>0.008</td>
<td>(0.247)</td>
<td>0.737**</td>
<td>(0.281)</td>
</tr>
<tr>
<td>Female</td>
<td>-0.483</td>
<td>(0.305)</td>
<td>-0.989**</td>
<td>(0.377)</td>
</tr>
<tr>
<td>Age</td>
<td>-0.335**</td>
<td>(0.064)</td>
<td>-0.054</td>
<td>(0.045)</td>
</tr>
<tr>
<td>Religiosity</td>
<td>-0.361**</td>
<td>(0.103)</td>
<td>0.050</td>
<td>(0.074)</td>
</tr>
<tr>
<td>Time</td>
<td>-0.019**</td>
<td>(0.006)</td>
<td>-0.006</td>
<td>(0.004)</td>
</tr>
<tr>
<td>Precision</td>
<td>-0.210**</td>
<td>(0.096)</td>
<td>0.577**</td>
<td>(0.173)</td>
</tr>
<tr>
<td>Improvement</td>
<td>0.081</td>
<td>(0.082)</td>
<td>0.258**</td>
<td>(0.071)</td>
</tr>
<tr>
<td>“High Bribe” Treatment</td>
<td>0.700**</td>
<td>(0.341)</td>
<td>-0.720</td>
<td>(0.321)</td>
</tr>
<tr>
<td>“High Wage” Treatment</td>
<td>-0.688</td>
<td>(0.351)</td>
<td>-0.829</td>
<td>(0.370)</td>
</tr>
<tr>
<td>“Low Monitoring” Treatment</td>
<td>-0.771</td>
<td>(0.366)</td>
<td>-0.462</td>
<td>(0.345)</td>
</tr>
<tr>
<td>“High Monitoring” Treatment</td>
<td>-0.263</td>
<td>(0.311)</td>
<td>-0.427</td>
<td>(0.364)</td>
</tr>
<tr>
<td>“No Bribe” Treatment</td>
<td>___</td>
<td>___</td>
<td>0.237</td>
<td>(0.369)</td>
</tr>
<tr>
<td>“Female * Monitoring”</td>
<td>___</td>
<td>___</td>
<td>1.664**</td>
<td>(0.623)</td>
</tr>
<tr>
<td>Accept</td>
<td>___</td>
<td>___</td>
<td>-0.527**</td>
<td>(0.230)</td>
</tr>
<tr>
<td>Ln L</td>
<td>-197.362</td>
<td></td>
<td>-129.138</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>204</td>
<td></td>
<td>247</td>
<td></td>
</tr>
</tbody>
</table>

Significance: * = 10%, ** = 5%, *** = 1%. The variable “Female * Monitoring” was not included in the Model “Rejecters” for lack of data.

† The endogenous variable is equal to 1 when the grader accepts the bribe.

* The endogenous variable is equal to 1 when the grader fails the briber.
<table>
<thead>
<tr>
<th>Variables</th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Grade Reported†</td>
<td>Grading Quality*</td>
</tr>
<tr>
<td>Age</td>
<td>-0.022 (0.030)</td>
<td>-0.136 (0.287)</td>
</tr>
<tr>
<td>Female</td>
<td>0.074 (0.270)</td>
<td>0.481 (1.627)</td>
</tr>
<tr>
<td>Religiosity</td>
<td>-0.036 (0.060)</td>
<td>-0.149 (0.463)</td>
</tr>
<tr>
<td>Time</td>
<td>-0.050 (0.039)</td>
<td>-0.368 (0.245)</td>
</tr>
<tr>
<td>Precision</td>
<td>0.858*** (0.110)</td>
<td>5.501*** (0.857)</td>
</tr>
<tr>
<td>Improvement</td>
<td>0.816*** (0.224)</td>
<td>4.830** (1.880)</td>
</tr>
<tr>
<td>Improvement * Exam Number</td>
<td>0.046* (0.019)</td>
<td>0.268** (0.113)</td>
</tr>
<tr>
<td>Reject * Control Treatment</td>
<td>-0.004 (0.267)</td>
<td>0.088 (2.184)</td>
</tr>
<tr>
<td>Reject * “High Bribe” Treatment</td>
<td>0.042 (0.380)</td>
<td>0.554 (3.770)</td>
</tr>
<tr>
<td>Reject * “High Wage” Treatment</td>
<td>0.339 (0.233)</td>
<td>2.268 (1.825)</td>
</tr>
<tr>
<td>Reject * “Low Monitoring” Treatment</td>
<td>-0.260 (0.285)</td>
<td>1.176 (1.769)</td>
</tr>
<tr>
<td>Reject * “High Monitoring” Treatment</td>
<td>-0.050 (0.227)</td>
<td>-0.638 (1.604)</td>
</tr>
<tr>
<td>Accept * Control Treatment</td>
<td>-0.739 (0.483)</td>
<td>-5.407*** (1.964)</td>
</tr>
<tr>
<td>Accept * “High Bribe” Treatment</td>
<td>-0.847* (0.444)</td>
<td>-7.544*** (2.281)</td>
</tr>
<tr>
<td>Accept * “High Wage” Treatment</td>
<td>-1.944 (1.489)</td>
<td>-13.572*** (3.319)</td>
</tr>
<tr>
<td>Accept * “Low Monitoring” Treatment</td>
<td>-0.926* (0.517)</td>
<td>-6.361*** (2.332)</td>
</tr>
<tr>
<td>Accept * “High Monitoring” Treatment</td>
<td>-0.181 (0.224)</td>
<td>-1.643 (1.729)</td>
</tr>
<tr>
<td>σ</td>
<td>1.920*** (0.048)</td>
<td>11.983*** (1.348)</td>
</tr>
<tr>
<td>σₐ</td>
<td>1.685** (0.527)</td>
<td>11.320*** (1.774)</td>
</tr>
<tr>
<td>σᵣ</td>
<td>0.902*** (0.087)</td>
<td>6.177*** (0.880)</td>
</tr>
<tr>
<td>Ln L</td>
<td>-5,459.470</td>
<td>-13,609.694</td>
</tr>
<tr>
<td>N</td>
<td>2,232</td>
<td>2,232</td>
</tr>
</tbody>
</table>

Significance: * = 10%, ** = 5%, *** = 1%. The model has been estimated with the data collected in all 6 treatments. The treatment effects are evaluated with respect to a grader of reference in the “No Bribe” treatment.

† The endogenous variable is $R_{it}$, the number of mistakes reported by grader $i$ for exam paper $t$ $(t=12,...,20)$.

* The endogenous variable is $-|R_{it} - M_i| / M_i$, where $M_i$ is the actual number of mistakes in exam paper $t$. 

The model has been estimated with the data collected in all 6 treatments. The treatment effects are evaluated with respect to a grader of reference in the “No Bribe” treatment.