Monitoring Corruption:
Evidence from a Field Experiment in Indonesia

Benjamin A. Olken
NBER
November 2004

ABSTRACT

This paper uses a randomized field experiment to examine several approaches to reducing corruption. I measure corruption in over 600 village road projects in Indonesia by having engineers independently estimate the prices and quantities of all inputs used in each road, and then comparing these estimates to villages’ official expenditure reports. I find that announcing an increased probability of a government audit, from a baseline of 4 percent to 100 percent, reduces theft from the project by about 8 percent of expenditures. I estimate that these audits are cost-effective. By contrast, I find that increasing grass-roots participation in the monitoring process reduced theft of villagers’ wages, but that this was almost entirely offset by corresponding increases in theft of materials. This suggests that grass-roots monitoring may be more effective in reducing theft when community members have substantial private stakes in the outcome, but less effective for public goods. Overall, the results suggest that traditional top-down monitoring can play an important role in reducing corruption, even in a highly corrupt environment.

*I wish to thank Alberto Alesina, Abhijit Banerjee, Robert Barro, Francesco Caselli, Joe Doyle, Esther Duflo, Amy Finkelstein, Brian Jacob, Seema Jayachandran, Ben Jones, Larry Katz, Michael Kremer, Jeff Liebman, Erzo Luttmer, Lant Pritchett, and Mark Rosenzweig for helpful comments. Special thanks are due to Victor Bottini, Richard Gnagey, Susan Wong, and especially to Scott Guggenheim for their support and assistance throughout the project. The field work and engineering survey would have been impossible without the dedication of Faray Muhammad and Suroso Yoso Oetomo, as well as the entire P4 field staff. This project was supported by a grant from the DFID-World Bank Strategic Poverty Partnership Trust Fund. All views expressed are those of the author, and do not necessarily reflect the opinions of DFID or the World Bank. Email: bolken@nber.org
1. Introduction

Corruption is endemic in much of the developing world. In many cases, corruption acts like a tax, adding to the cost of providing public services and conducting business. Often, though, the efficiency costs of corruption can be far worse.\(^1\) Indeed, it has been suggested that corruption may be a major contributor to the low growth rates of many developing countries (Mauro 1995).

Despite the importance of the problem, the inherent difficulty of directly measuring corrupt activity has meant that there is relatively little evidence, and therefore relatively little consensus, on how to best reduce corruption. One approach to reducing corruption, dating back at least to Becker and Stigler (1974), suggests that the right combination of monitoring and punishments can control corruption. In practice, however, the very individuals tasked with doing the monitoring and enforcing the punishments may themselves be corruptible. In that case, increasing the probability that a low-level official is monitored by a higher-level official could result only in a transfer between the officials, not in a reduction of corruption.\(^2\) Whether one can actually control corruption by increasing monitoring in such an environment is an open, and important, empirical question.

An alternative approach to reducing corruption, which has gained prominence in recent years, is to increase grass-roots participation by community members. Community participation is now regarded in much of the development community as the key not only to reduced corruption, but to improved public service delivery more generally. For example, the entire 2004 World Development Report is devoted to the idea of “putting poor people at the center of service

---

\(^1\) See, for example, Krueger (1974) and Shleifer and Vishny (1993) for examples of how the efficiency costs of corruption can substantially exceed the amount stolen itself.

\(^2\) Cadot (1987), for example, discusses this possibility, and shows that this type of multi-tiered corruption can lead to multiple equilibria in corruption.
provision: enabling them to monitor and discipline service providers, amplifying their voice in policymaking, and strengthening the incentives for service providers to serve the poor.” (World Bank 2004) The idea behind the grass-roots approach is that community members are the people who benefit from a successful program, and so may have better incentives to monitor than disinterested central government bureaucrats (Stiglitz 2002). Of course, this approach has potential drawbacks as well – for example, monitoring public projects is a public good, so there may be a serious free rider problem. Grass-roots monitoring may also be prone to capture by local elites (Bardhan 2002, Bardhan and Mookherjee 2003). Given these countervailing forces, whether grass-roots monitoring can actually succeed in reducing corruption is also an empirical question.3

To examine these alternative approaches to fighting corruption, I designed and conducted a randomized, controlled field experiment in 608 Indonesian villages. At the time the study started, each village in the study was about to start building a village road as part of a nationwide village-level infrastructure project. To examine the impact of external monitoring, I selected some villages to be told, after funds had been awarded but before construction began, that their project would subsequently be audited by the central government audit agency. This amounted to increasing the probability of an external government audit in those villages from a baseline of about 4 percent to essentially 100 percent. Government audits carry with them the theoretical possibility of criminal action, though this is quite rare; more important, the results of the audits were read publicly to an open village meeting by the auditors, and so could result in substantial social sanctions. The audits were subsequently conducted as promised.

3 Several authors have found suggestive evidence in both micro and macro cross-sectional data that higher levels of “voice” are associated with lower levels of corruption. Rose-Ackerman (2004) provides a summary of much of the work on this topic to date.
To investigate the impact of increasing community participation in the monitoring process, I designed two different experiments that sought to increase grass-roots monitoring in the project. Specifically, the experiments sought to enhance participation at “accountability meetings,” the village-level meetings where project officials account for how they spent project funds. In one experiment, hundreds of invitations to these meetings were distributed throughout the village, to encourage direct participation in the monitoring process and to reduce elite dominance of the process. In the second experiment, an anonymous comment form was distributed along with the invitations, providing villagers an opportunity to relay information about the project without fear of retaliation. This comment form was then collected before the accountability meetings in sealed drop-boxes, and the results were summarized at the meetings. Both of these experimental interventions were successful in raising grass-roots participation levels—the invitations increased the number of people participating in the accountability meetings by about 30 percent, and the comment forms generated hundreds of comments about the project, both good and bad, in each village.

To evaluate the impact of these experiments on corruption, one needs a measure of corruption. Traditionally, much of the empirical work on corruption has been based on perceptions of corruption, rather than on direct measures of corruption. This paper, however, builds on a small but growing literature that measures corruption directly, by comparing two measures of the same quantity, one “before” and one “after” corruption has taken place. To

---

4 The use of perceptions-based measures of corruption in economics was pioneered by Mauro (1995) and forms the basis of the much-cited Transparency International Corruption Index (Lambsdorff 2003). More recent work using perceptions-based measures is summarized in Rose-Ackerman (2004).

5 For example, Reinikka and Svensson (2004) examine corruption in educational expenditures, Fisman and Wei (forthcoming) and Yang (2004) examine corruption in international trade, Di Tella and Schargrodsky (2003) examine corruption in hospital procurement, and Olken (2004) examines theft from a government redistribution program. This paper slightly from much of this literature by comparing government reports to an independently constructed estimate, rather than comparing government reports to the reports of a different government agency or to a household survey.
measure corruption in the road projects, I assembled a team of engineers and surveyors who, after the projects were completed, dug core samples in each road to estimate the quantity of materials used, surveyed local suppliers to estimate prices, and interviewed villagers to determine the wages paid on the project. From these data, I construct an independent estimate of the amount each project actually cost to build, and then compare this estimate with what the village reported it spent on the project on a line-item by line-item basis. Since the village must account for every Rupiah it received from the central government, stolen funds must show up somewhere in the difference between reported expenditures and estimated actual expenditures.

Using these data, I find that there were substantial reductions in corruption associated with the audit experiment. In particular, I estimate that the audit treatment – i.e., increasing the probability of an audit from a baseline of 4 percent to 100 percent – was associated with reductions in corruption of an average of 8 percent of expenditures. These reductions came both from reductions in over-invoicing of materials procured for the project and in reductions in over-billing of labor. Interestingly, I find that nepotism – i.e., the number of project jobs given to family members of project officials – actually increased in response to the audits. I compare the costs and benefits from the audits and find that the audits were highly cost effective.

By contrast, I find that the participation experiments – the invitations and the comment forms – altered the method of corruption chosen by village officials, but had little impact on overall corruption levels. As discussed above, the interventions did raise community participation in the monitoring process. Moreover, villages in the invitations treatment were more likely to openly discuss corruption problems at the accountability meetings, and villages receiving both invitations and comment forms were more likely to take serious action at the meeting to resolve corruption-related problems. However, the magnitude of these changes in
behavior at the meetings was small, and these treatments did not measurably reduce overall corruption in the projects.

What these treatments did do, however, is to change the way in which corruption was hidden. In particular, I find that, relative to control villages, villages in the invitations treatment chose to hide less corruption in overstated wage payments and, instead, to hide more corruption in overstated expenditures on materials. The results suggest that the decision of how to allocate actual expenditures and where to hide corruption are, to a large degree, separable decisions. The results suggest that the change in where corruption was hidden was driven by increased participation in accountability meetings by project workers, who have a personal stake in ensuring that their wages are not stolen. This suggests that grass-roots monitoring may be most effective when a subset of people personally stand to gain from reducing corruption, and therefore have strong incentives to monitor actively. Grass-roots monitoring may be less effective for projects that provide public goods, such as the infrastructure projects studied here.

The remainder of the paper is organized as follows. Section 2 discusses the setting in which the study takes place. Section 3 discusses the conceptual issues examined in the paper. Section 4 describes the experimental interventions. Section 5 describes the data used in the study. Section 6 presents the results of the experiments. Section 7 performs a cost-benefit analysis. Section 8 concludes.

2. Setting

The Kecamatan (Subdistrict) Development Project, or KDP, is a national government program, funded through a loan from the World Bank. KDP finances projects in approximately 15,000 villages throughout Indonesia each year. The data in this paper come from KDP projects in 608 villages in two of Indonesia’s most populous provinces, East Java and Central Java.
In KDP, participating subdistricts, which typically contain between 10 and 20 villages, receive an annual block grant for three consecutive years. Every year, each village in the subdistrict makes a proposal for any combination of small-scale infrastructure and seed capital for microcredit cooperatives. The majority of villages (72%) propose an infrastructure project, plus a small amount for savings-and-loans for women. An inter-village forum ranks all of proposals according to a number of criteria, such as number of beneficiaries and project cost, and projects are funded according to the rank list until all funds have been exhausted. Villages that receive funding hold a planning meeting to determine how to implement the project, and then proceed with construction over the subsequent three-to-four months.

A typical funded village receives funds on the order of Rp. 80 million (US$8,800) for infrastructure; these funds are often supplemented by voluntary contributions from village residents, primarily in the form of unpaid labor. These projects are large relative to ordinary local government activities; in 2001, the average village budget was only Rp. 71 million (US$7,800), so receiving a KDP project more than doubles average local government expenditures. The allocation to the village is lump-sum, so that the village is the residual claimant. In particular, surplus funds can be used, with the approval of a village meeting, for additional development projects, rather than having to be returned to the KDP program.

By far the most common type of infrastructure project proposed by villages is the surfacing of an existing dirt road with a surface made of sand, rocks, and gravel. These roads range in length from 0.5 – 3 km, and may either run within the village or run from the village to the fields. Dirt roads in Java are typically impassible during the rainy season; surfacing these roads allows them to be used year-round. To facilitate comparisons, the sample of villages considered in this paper is limited to villages with such non-asphalt road projects.
A number of different actors play a role in the KDP process. At the beginning of the project, each village elects 5 people to administer the project in the village—two “facilitators,” whose main role is to help solicit suggestions for what project to build and to facilitate village meetings, and three members of an implementation team—a chairman, secretary, and treasurer—who are actually responsible for building the project. Once the project is funded, money is disbursed directly to the implementation team. The team procures materials and hires labor directly, rather than using contractors. The members of the implementation team receive an honorarium, limited in total to a maximum of 3 percent of the total cost of the project. In addition to the village facilitators and implementation team, there is technical and administrative support provided by two full-time, government-contracted subdistrict facilitators, as well as supervision at the district, provincial, and national levels. Village government members, particularly the village head, have no official role in the project, but are often actively involved on an informal basis.

The project includes several mechanisms to ensure proper use of project funds. The primary mechanism is a series of village-level accountability meetings. Funds are released to the implementation team in three tranches, of 40%, 40%, and 20% of the funds, respectively. In order to obtain the second and third tranches of funds, the implementation team is required to present an accountability report to an open village meeting, explaining how all funds were used. Only after that meeting has approved the accountability report is the next tranche of funds released. Similarly, in order to participate in the subsequent year of KDP, villages are required to present a final, cumulative accountability report at the end of the project, which similarly must be approved by a village meeting. Though open to the entire village, these meetings are typically

---

6 In the data, the total honorarium payments to each member of the implementation team averaged Rp. 460,000, or approximately Rp. 150,000 ($17) per month of the construction. By comparison, median per-capita monthly expenditure in comparable areas in East and Central Java in 2003 was about Rp 140,000 ($15).
attended by only 30-50 people, most of whom are members of the village elite, out of a total adult population of about 2,500.

In addition, district and provincial level project managers conduct overall supervision of all projects, and there is a provincial complaints-handling unit which investigates allegations of improprieties. Furthermore, each year, the project is audited by the independent government development audit agency, *Badan Pengawasan Keuangan dan Pembangunan* (BPKP). Each village-level project in the study area has about a 4% baseline chance of being audited by BPKP. Auditors cross-check all of the financial records looking for irregularities, as well as inspect the physical infrastructure. Findings from the audits are sent to project officials for follow-up, and can potentially lead to criminal action, though prosecutions for village-level officials are in practice quite rare. (Woodhouse 2004) More often, officials found to have stolen money are forced to publicly return the money, which can result in substantial social sanctions.

Corruption at the village level can occur in several ways. First, implementation teams, potentially working with the village head, may collude with suppliers. Suppliers can inflate either the prices or the quantities listed on the official receipts to generate money for a kick-back to village and project officials. Second, members of the implementation team may manipulate wage payments. Frequently, this is accomplished by convincing villagers to accept a lower-than-market wage—and in some cases to work completely for free. While in many cases this voluntary labor is actually used to extend the road or to build another ancillary project, corrupt officials can instead not report the voluntary contributions and simply pocket the difference. In other cases, those running the project can simply inflate the number of workers paid by the project. All of these types of corruption will be investigated in the empirical work below.

---

7 Of course, there may also be collusion or kickbacks at the national or district level of the program. This paper, however, focuses on corruption where the bulk of the program money is actually spent—at the village level.
3. Conceptual Approach

This section introduces a simple framework to examine the decision faced by potentially corrupt officials. Going back at least to Becker and Stigler (1974), economists have posited that officials balance the gains from corruption with the risk of punishment, and choose corruption if the gains from being corrupt exceed the expected punishment. The framework used here is based on the ideas of Becker and Stigler, but adapted to allow for multiple types of corruption that can be differentially affected by different types of monitoring.8

To fix ideas, suppose that there are two methods through which an official can steal money from a project, denoted type 1 and type 2. Denote by $c_1$ the fraction of the total value of the project stolen through method 1, and by $c_2$ the fraction of the total value of the project stolen through method 2.

If the official is caught stealing, he must pay a fine $F$.9 The probability that the official is caught stealing is a convex function of the amount of corruption stolen through each method, which I parameterize as follows:

$$P(c_1, c_2) = \alpha \left( \gamma_1 c_1^2 + \gamma_2 c_2^2 + \delta c_1 c_2 \right)$$

The official therefore solves the following expression:

$$\max_{c_1, c_2} \quad c_1 + c_2 - F \alpha (\gamma_1 c_1^2 + \gamma_2 c_2^2 + \delta c_1 c_2)$$

(1)

As is clear from expression (1), the parameter $\delta$ captures the degree to which method 1 and method 2 are complements or substitutes—if $\delta > 0$, then method 1 and method 2 are substitutes, and if $\delta < 0$, method 1 and method 2 are complements.

---

8 The framework presented here is also related to the one presented by Yang (2004). The main difference is that Yang adds fixed costs to using the second method of corruption, and investigates the impact of partial audits on switching between methods, whereas I consider the case where the official uses both methods simultaneously and there are no fixed costs.
9 In practice, of course, this fine need not be monetary; social sanctions may be as effective as monetary fines, if not more so.
This simple setup illustrates several issues. First, under the assumptions that guarantee an interior solution in $c_1$ and $c_2$ (namely, that for both methods of corruption $j$, $\gamma_j > |\delta|$), it follows directly from expression (1) that both types of corruption, and therefore total corruption, are decreasing in the probability of detection ($\alpha$) and in the punishment ($F$).

Second, in practice, different types of monitoring behavior are more successful at detecting some methods of corruption than others. For example, in Indonesia, the audits conducted by BPKP checked the administrative records for price markups and forged receipts, and checked the physical infrastructure for dramatic reductions in the visible components of the road. On the other hand, the audits did not directly inspect the layer of sand underneath the rocks, and did not investigate who the workers on the project were. Similarly, villagers might be much better at detecting whether wage rates paid to village laborers were too low than whether the quantity of materials delivered to the project was correct.

In the language of the framework, an increase of monitoring of one type of corruption can be captured by an increase in $\gamma_1$. One can show that, under the same assumptions guaranteeing an interior solution, while increasing the monitoring on method 1 ($\gamma_1$) unambiguously decreases both method-1 corruption ($c_1$) and the total amount of corruption ($c_1 + c_2$), the impact on method 2 corruption ($c_2$) is theoretically ambiguous, and depends on whether the method 1 and method 2 are substitutes for or complements with one another.\(^\text{10}\) If the efficiency costs from method 2 are greater than those from method 1, and method 1 and method 2 are sufficiently strong substitutes, then it is possible that partial audits could increase the total efficiency losses from corruption.

In the setup above the level of monitoring was treated as exogenous. But in the case of grass-roots monitoring, community members chose both how much to monitor ($\alpha$) and how to

\(^{10}\) As shown by Yang (2004), in fact, in the presence of fixed costs, total corruption can actually increase in response to a partial increase in monitoring.
allocate their monitoring among types of corruption ($\gamma_1$ and $\gamma_2$). In that case, the amount of
monitoring chosen will depend both on the costs of monitoring each type of corruption, and the
benefits each individual personally gains from monitoring. The people who may be most willing
to incur these costs are those who stand personally to gain from reducing corruption. For
example, those who work on the project, and therefore receive wages from the project in addition
to enjoying the benefits of the road, have a stronger incentive to monitor than non-workers. They
will also prefer a different composition of monitoring activity than non-workers, as they will be
more likely to focus on theft of their wages than on other types of theft. Thus, the differences in
monitoring ($\gamma_1$ and $\gamma_2$) may represent differences in monitoring intensity, as well as differences in
the ease with which different types of corruption can be detected.

Finally, so far it was assumed that there were honest external enforcers, who charge a
fine $F$ if and only if they detect corruption. But what would happen if the enforcers were
themselves briable?11 The amount of the bribe (if it was paid at all) would then be the outcome
of a bargaining game between the enforcers and the officials, which would likely depend on the
enforcers’ threat point. If the enforcers can impose the fine $F$ regardless of whether they find
corruption, the expected bribe will not depend on the amount of corruption in the project.
Anticipating this, an increase in the probability of an external audit reduces the expected income
of the corrupt officials, but (assuming no income effects) it does not change their marginal
incentives to steal, and thus will not affect the amount of corruption they chose. Alternatively, if
the enforcers can impose the fine $F$ only if they detect corruption, and if the probability of
detection is increasing in the amount of corruption (as above), then the expected bribe will be

---

11 There is reason to think they may be. According to Richard Gnagey, the chief engineer of KDP, “draft audit
reports are filled with things that disappear before the final report. And many things are negotiated away even before
appearing in the draft.” (Personal correspondence, September 30, 2004).
increasing in the amount of corruption. In that case, increasing the probability of an external audit can still reduce corruption, even if the auditors are corrupt.

4. Experimental Design

The experiments discussed in this paper examine different ways of altering the probability that corruption is detected and punishments are enforced. Three interventions are examined – increasing the probability of external audits (“Audits”), increasing participation in accountability meetings (“Invitations”), and providing an anonymous comment form to villages (“Comments”). Section 4.1 discusses the overall experimental design. Section 4.2 then discusses the Audit interventions, and Section 4.3 describes the Invitations and Comment interventions.

4.1. Experimental Design

Table 1 displays the basic experimental design. As shown in Table 1, randomization into the Invitations and Comment Form treatments was independent of randomization into the Audit treatment. In both cases, the treatments were announced to villages after the project design and allocations to each village had been finalized, but before construction or procurement of materials began. Thus, the choice of what type of project to build, as well as the project’s design and planned budget, should all be viewed as exogenous with respect to the experiments.

The randomization of Audits and Invitations/Comment treatments differed in several ways. First, the treatments differed in the degree to which they were expected to have spillover effects in surrounding villages. For Audits, the concern was that if one village was told that they would be audited, officials in other villages might worry that when the auditors came to the

---

12 In all villages (including control villages), at the village meeting immediately after the final allocations were announced, the study enumerator made a short (< 5 minute) presentation, introducing him or herself and explaining that there would be a study in the village, that each village and project official would be interviewed for data collection, and that the enumerator would be present to record what happened at each of the accountability meetings. In villages receiving a treatment, the only difference was that this introduction was followed by a description of the treatment(s) in that village. The final engineering survey was not mentioned at all to the villagers during this presentation, or subsequently, until the surveyors actually appeared to conduct the survey.
subdistrict, their villages might be audited as well. If these surrounding villages were used as controls, this would understate the effects of the audits. On the other hand, the participation treatments were much less likely to have similar spillover effects, as the treatment was directly observable in the different villages early on. Therefore, the randomization for Audits was clustered by subdistrict (i.e., either all study villages in a subdistrict received Audits or none did), while the randomization for Invitations and Comment Forms was done village-by-village. The calculations of the standard errors below are adjusted to take into account the potential correlation of outcomes in villages within a subdistrict.

This difference in clustering also necessitated a difference in stratification. As the Invitations and Comment Forms were randomized village-by-village, they were stratified by subdistrict, the lowest administrative level above the village. Since the Audits were randomized by subdistrict, they needed to be stratified at a higher level. Therefore, the Audits were stratified by district and by the number of years the subdistrict had participated in the KDP program. This yielded a total of 156 strata for the Invitations / Comment Forms, each containing an average of 3.8 study villages, and 50 strata for the Audits, each containing an average of 3.1 study subdistricts and 12.1 study villages.

In the analysis, I report three specifications – no fixed effects, fixed effects for each engineering team that conducted the survey, and stratum fixed effects. Despite the stratification, the randomization was designed so that the probability each village received a given treatment was always held constant, regardless of what stratum the village was in. The probability of receiving a given treatment is therefore orthogonal to any stratum or village-level variable, so

---

13 This was most likely to be a problem within subdistricts, as there is frequent communication between both village officials and project officials within a subdistrict. Communication across subdistrict lines is much more limited, particularly for village officials.
including stratum fixed effects, while it may reduce standard errors, is not necessary for the analysis to be consistent.\textsuperscript{14}

4.2. \textit{The Audit Experiment}

In the “Audit” treatment, villages were told, at the village meeting where they began planning for actual construction, that their project would be audited by BPKP, the government audit agency, with probability 1. They were told that the audit could take place either during or after construction was finished, and would include inspections of both the project’s financial records as well as a field-inspection of the construction activities. Approximately two months later, the village implementation team received a one-page letter from BPKP which confirmed that the village had been chosen to be audited and which spelled out in somewhat greater detail exactly what would be covered by the audit.

Villages were told that results of the audits, in addition to being reported to the central government and project officials, would also be delivered directly by the auditors to a special village meeting. Village officials therefore faced several potential sanctions from the audits – retribution from the village, the possibility that the village would not receive KDP projects in the future, and the theoretical possibility of criminal action, though the latter is quite rare.

Between one and four months after the projects had started, Phase I of the audits commenced.\textsuperscript{15} The main purpose of this first round of audits was to credibly demonstrate that the

\textsuperscript{14} In this case, however, there is a disadvantage to including stratum fixed effects. In particular, the audit randomization was conducted before the list of villages with road projects was known (though the randomization results were kept strictly secret). Out of the initial 166 subdistricts included in the randomization, only 156 subdistricts ended up having villages with road projects and could be included in the study. As a result, 7 out of 50 strata contain either all audit or all non-audit villages. Thus, while including stratum fixed effects reduces the variance of the error term, and thus improves statistical power, it also reduces the effective sample size, as these 7 strata can no longer be used. Note also that, since each stratum was surveyed by only one engineering team, stratum fixed effects implicitly capture engineer fixed effects.

\textsuperscript{15} All audits in this phase took place during a three week period during mid January / early February 2004. However, since there was heterogeneity in the timing of when construction started, this was anywhere between 1 and 4 months after construction had begun.
audits were real, rather than an idle threat. One village in each subdistrict receiving the Audit treatment was randomly selected to be audited during this first phase. The audit was conducted over two days, and the results were subsequently presented by the auditors to a specially called village meeting. The meeting was open to the public, and was attended by an average of 44 village members – similar to the number who attended other project meetings in the village. At that meeting, the results of the audit were read, and members of the implementation team and village officials were given an opportunity to propose corrective actions for the auditors’ findings.

After the first round of audits, all study villages receiving the audit treatment, including the village that was audited during Phase I, were informed in another letter from BPKP that they would be audited again after the construction on the project had been completed. The second phase of the audits was conducted approximately seven months subsequently, after both construction was finished and the collection of the corruption data described below was completed. As with Phase I audits, the results from the Phase II audits were presented to the village in an open village meeting and forwarded to the project for followup.

4.3. The Participation Experiments

In the “Invitations” treatment, either 300 or 500 invitations were distributed throughout the village several days prior to each of the three accountability meetings.\(^\text{16}\) Though village meetings are officially open to the public, in practice Javanese villagers consider it quite rude to attend a meeting to which they have not been formally invited (usually in writing), and with the exception of a few independent-minded members of the village elite, they rarely do. The village head, who normally issues written invitations for the meetings, therefore has the potential to

\(^{16}\) In addition, for each meeting a small subsidy – Rp. 45,000 ($5) for villages with 300 invitations, Rp. 75,000 ($8) for villages with 500 invitations – was given to the implementation team to cover the additional cost of providing snacks to the extra attendees induced by the invitations.
stack the attendance of the accountability meeting in his favor by issuing invitations only to his supporters. By distributing a large number of invitations, the village head’s ability to control who attends the meeting was substantially reduced.

Given the size of a typical village, approximately one in every two households in treatment villages received an invitation. The invitations were distributed either by sending them home with school children, or by asking the heads of hamlets and neighborhood associations to distribute them throughout their areas of the village. The number of invitations (300 or 500) and the method of distributing them (schools or neighborhood associations) was randomized by village.17

In the “Invitations + Comment Forms” treatment, invitations were distributed exactly as in the Invitations treatment. However, attached to the invitation was a comment form asking villagers’ opinions of the project. The form asked the recipient to answer several questions about the road project, and then to return the form – either filled out or blank – to a sealed drop box, placed either at a village school or at a store in the sub-village. The instructions stated clearly that the recipients should not write their name on the form, in order to preserve their anonymity. The form had four closed-response questions (i.e., requesting answers of the form Good, Satisfactory, or Poor), asking recipients about their knowledge of the project, opinion of the project’s financial management, opinion of the project’s physical quality, and opinion of the prices being paid by the project. In addition, the form had two free-response questions, one asking respondents to evaluate the job performance of the implementation team and one asking respondents about any other project-related issues they felt needed to be discussed.

17 The purpose of these extra randomizations – the number of invitations and how they were distributed – was to generate additional variation in the number and composition of meeting attendees, to distinguish size effects from composition effects. However, as shown in Table 5 below, these sub-treatments did not generate a strong enough first-stage to be used in the analysis.
The comment forms were collected from the drop boxes two days before each meeting and summarized by a project enumerator. The summary was then read by the enumerator at the meeting. For the closed-response questions, the summary consisted simply of the number of responses in each category. For the open-response questions, the enumerator categorized each comment into one of 11 possible response categories. The summary consisted of the number of comments in each category, along with a representative sample of actual comments from the five most common categories of responses.18

5. Data

The data used in this paper come from four types of surveys, each designed by the author and conducted specifically as part of the project: a key-informant survey, covering baseline characteristics about the village and the village implementation team; a meeting survey, containing data on the attendees and a first-hand report of discussions at the accountability meetings; a household survey, containing data on household participation in and perceptions of the project; and a final field survey, used to measure corruption in the project. This measurement was conducted in all villages (both treatment and control), and is completely separate the audits conducted by BPKP as part of the “Audit” treatment. This section describes the final field survey used to measure corruption; the remaining data, as well as additional details on the field survey, are discussed in more detail in Appendix A.

5.1. Reported Expenditures

The basic strategy I use to measure corruption is to compare what villages claim they spent on the project to an independent estimate of what villages actually spent. Obtaining data on

---

18 To minimize the possibility that the comment form process could be used for isolated vendettas by disgruntled individuals in the village, enumerators were instructed that comments read aloud should not mention anyone by name or by position unless there were at least five essentially similar comments about the person.
what villages claim they spent is relatively straightforward. At the end of the project, all village implementation teams are required to file an accountability report with the project subdistrict office, in which they report the prices, quantities, and total expenditure on each type of material and each type of labor (skilled, unskilled, and foreman) used in the project. The total amount reported must match the total amount allocated to the village. In addition, they also report, for each type of material and labor, the amount donated to the project by villagers. These financial reports were readily available to the survey team for all study villages.

Obtaining an independent estimate of what was actually spent is substantially more difficult, and involves three main activities—an engineering survey to determine quantities of materials used, a worker survey to determine wages paid by the project, and a supplier survey to determine prices for materials.

5.2. Measuring Quantities of Materials

In the engineering survey, an engineer and an assistant conducted a detailed physical assessment of all physical infrastructure built by the project in order to obtain an estimate of the quantity of materials used. In the standard road design, known as a Telford road, the road consists of three types of materials—a base of sand, a layer of large (10-15 cm) rocks, and a top layer of gravel to provide a smooth running surface.\(^{19}\) To estimate the quantity of each of these materials, the engineers dug ten 40cm × 40cm core samples at randomly selected locations on the road. By combining the measurements of the volume of each material per square meter of road with measurements of the total length and average width of the road, I can estimate the total quantity of materials used in the road.

\(^{19}\) Three other similar road designs are also included in the study. Telasah roads are similar to Telford, but install the rocks flat-side-up to create a smooth running surface, and therefore largely omit the gravel layer. Sirtu roads consist of gravel only, omitting the sand and rock layer. Katel roads are similar in design to Telford, but use a mixture of clay and gravel in the top layer to create a more permanent top surface. Telford roads, however, account for 86 percent of the road projects in the sample.
It is important to note, however, that this estimate of the materials used in the road, while it should be proportional to the total quantity of materials used in the road, may be smaller in magnitude than the actual amount of materials used in the road, as some amount of loss is normal during construction and measurement. For example, some amount of sand may blow away off the top of a truck, or may not be totally scooped out of the hole dug by the engineers conducting the core sample. I denote the average percentage of materials lost due to normal construction processes and measurement error but with no corruption as the \textit{loss ratio}.

To deal with these loss ratios, whenever possible I express the measured quantities in log form, so that the average loss ratio will be captured by the constant term and will not affect estimated differences across villages. However, in some cases, such as when combining different types of materials into an aggregate measure of corruption, this approach is not sufficient, and one actually needs to estimate these loss ratios. One also needs an estimate of these loss ratios if one is interested in the level of corruption, not just the differences across villages.

To obtain such an estimate, I constructed four short (60m) “test roads” in different areas of East and Central Java as a calibration exercise. During the construction of each of these roads, the survey team carefully measured all quantities before construction (i.e., while still in the delivery trucks). Materials were again measured after construction was completed using the techniques described above. The ratio between the amount of materials actually used in the road and the amount measured after the road was built is an estimate of the loss ratio. I describe this calibration exercise and the resulting loss ratios in more detail in Appendix B.

While the road project comprises the main use of KDP funds in each village, roads are often accompanied by smaller ancillary projects, such as culverts, retaining walls, and gabions, and occasionally by larger projects, such as a small bridge. For each of the ancillary projects, the
engineer on the survey team conducted a detailed field survey, measuring and sketching each constructed piece of infrastructure to estimate the volume of materials, such as cement, rocks, and sand, used in the construction.

5.3. *Measuring Wages and Hours Worked*

Workers, defined as people who worked on the project for pay, were asked which of the many activities involved in building the road were done with paid labor, voluntary labor, or some combination, what the daily wage and number of hours worked was, and to describe any piece rate arrangements that may have been part of the building of the project. To estimate the quantity of person-days actually paid out by the project, I combine information from the worker survey about the percentage of each task done with paid labor, information from the engineering survey about the quantity of each task, and assumptions of worker capacity derived both from the experience of field engineers and the experience from building the test roads. These assumptions of worker capacity are discussed in more detail in Appendix B.

5.4. *Measuring Prices*

Since there is substantial variation in the prices of construction materials across subdistricts, a price survey was conducted in each subdistrict. Since there can be substantial differences in transportation costs within a subdistrict, surveyors obtained prices for each material that included transportation costs to each survey village. The price survey included several types of suppliers—supply contractors, construction supply stores, truck drivers (who typically transport the materials used in the project), and workers at quarries—as well recent buyers of material (primarily workers at construction sites).\(^2\) For each type of material used by

\(^2\) Furthermore, to reduce the potential for bias induced by surveying the actual suppliers for the project, who may be in collusion with project officials, only survey responses from sources outside a given village are used to construct the prices for that village, and no mention of KDP was made until the end of the interview. It turns out that 27% of those interviewed for the price survey had actually been suppliers to the KDP program; dropping them from the
the project, between three and five independent prices were obtained; I use the median price from the survey for the analysis.

5.5. Measures of Corruption

From these data, I can calculate the expenditures from the financial reports, which I hereafter will refer to as reported amount, and the expenditures estimated from the field survey, which I will refer to as the actual amount. I define the percent missing to be the difference between the log of the reported amount and the log of the actual amount. This variable—the percent missing—is the main measure of corruption used in the subsequent analysis.

I use several different versions of the percent missing measure in the empirical analysis. The primary measure is the percent missing for the four major items—sand, rocks, gravel, and unskilled labor—used in the road project. As shown in Table 2, expenditures on these four items account for 88% of reported expenditures on the road project. As these are the four major sources of expenditure, substantial effort was put into ensuring that these four items were measured as accurately as possible in the engineering survey.

I report several other corruption measures as well. As shown in Table 2, the road project accounts for 76% of total funds spent; a further 15% of funds are spent on ancillary projects that go along with the road, such as culverts and retaining walls. Each of these projects was inspected by the field engineers, generating an estimate of the amount of sand, rock, cement, and labor used in each. The second measure of corruption, major items in the main road + ancillary projects, adds in these expenditures as well.21 Finally, I report the percent missing separately for materials and unskilled labor.

analysis, however, does not affect the results. In fact, restricting the price data to only prices obtained from buyers of materials (i.e., dropping all suppliers of materials from the price survey) also does not affect the results.

21 The number of observations is higher when these ancillary expenditures are included because, for some villages, reported expenditures for the main road were combined with those from the ancillary projects. For those villages, the
6. Experimental Results

6.1. Estimating Equation

Given the randomized nature of the experiments, estimating their effects is straightforward. I estimate an equation of the following form via OLS:

\[
\text{PERCENTMISSING}_{ijk} = \alpha_1 + \alpha_2 \text{AUDIT}_{jk} + \beta_3 \text{INVITATIONS}_{ijk} + \\
\beta_4 \text{INVITATIONSANDCOMMENTS}_{ijk} + \epsilon_{ijk}
\]  

(2)

where \(i\) represents a village, \(j\) represents a subdistrict, and \(k\) represents a stratum for the audits.

Since the AUDIT treatment variable is perfectly correlated within subdistricts, the standard errors are adjusted to allow for correlation within subdistricts. As each of the 12 engineering teams may have conducted the corruption measurements slightly differently, I estimate a version of equation (2) that includes engineering team fixed effects. Finally, when investigating the audits, I estimate a version of equation (2) that includes fixed effects for each audit stratum \(k\), and when investigating the invitations and comment forms, I estimate a version of equation (2) that includes fixed-effects for each subdistrict \(j\) (i.e., the stratifying variable for the participation experiments).\(^{22}\)

6.2. The Audit Experiment

6.2.1. Overall Effects on Corruption

Table 3 presents the main results from the audit experiment. Each row presents the percent missing in different aspects of the project. The first column presents the mean percent

\(^{22}\) Note that approximately 10% of the observations in the sample were dropped, because the reported expenditure could not be accurately matched to the data from the engineering survey. This was caused by one of four reasons: (1) surveyor error in measuring the road, (2) the project consisted largely of a partial rehabilitation of an existing road, (3) agglomerated expenditures reports (i.e., the village expenditure report combined expenditures in the road project with other projects that could not be independently measured, such as a school), or (4) villages that had asphalted the road that refused to let the engineers break the asphalt to conduct the engineering survey. The probability of an observation being dropped for any one of these reasons appears empirically to be completely orthogonal to any of the treatments.
missing in the control villages – i.e., those villages that did not receive the audits – and the second column presents the mean level in the villages that received the audits. The effect of the audits – i.e., the coefficient \( \alpha_2 \) in equation (2) – is presented in column (3). The p-value from a test that the audit effect is zero is presented in column (4). Columns (5) and (6) again present the audit effect and p-values allowing for engineer fixed effects, and columns (7) and (8) present the results allowing for stratum fixed effects.

The results show that the audits had a substantial, and statistically significant, negative effect on theft from the project. For the major items in the road project, the point estimate is that the audits reduced the percent missing by between 5.6 and 9.0 percentage points, depending on the specification. When ancillary items are also included, the point estimate is a reduction in the percent missing of between 7.2 and 7.9 percentage points. Depending on the specification, these effects are generally statistically significant at between the 0.05 and 0.10 significance levels. Taken as a whole, I conclude that the audits reduced corruption by about 8 percentage points of the cost of the project. Compared with the baseline level of about 30 percentage points in control villages, the point estimates imply a reduction in theft of between 20 and 25 percent of the baseline level, although, as discussed above, the absolute levels of corruption depend on assumptions for loss ratios and should be interpreted with some caution.

Breaking down the change in corruption into materials and labor, there appears to be substantial reductions in corruption in both materials (rocks and gravel) and unskilled labor, though these separate effects are not statistically significant.\(^{23}\)

Mechanically, the theft I measure must occur either by inflating the price charged per unit or by overstating the actual quantities used. When the goods being procured are commodities (as

---

\(^{23}\) It is worth noting that, due to the log transformation, mechanically the change in \( \log(\text{materials} + \text{wages}) \) and \( \log(\text{materials}) + \log(\text{wages}) \) will not be identical.
they are in this case), it is much easier to verify the unit price than to verify the quantity of materials used. Consistent with this, when I decompose the results into differences in prices and differences in quantities, the point-estimates suggest almost no corruption on the price dimension and show that all the corruption occurred on the quantity dimension.\textsuperscript{24} Consistent with there being more theft on the quantity margin rather than on the price margin, I also find that the reductions in corruption due to the audits come primarily through quantities, particularly for rocks, though these breakdowns are generally not statistically significant.

One potential concern is whether the observed effects actually represent a change in corruption in response to the audit treatment, or whether incompetent builders are simply replaced by more skilled builders—i.e., is the corruption measure picking up corruption or competence? To investigate this, I examine a number of quality measures, such as compactness of the road, the size and shape of the rocks, and the grade of the road, all of which are relatively inexpensive. Overall competence should affect both expensive and inexpensive components, whereas a reduction in corruption would affect only the expensive aspects of construction (i.e. volume of materials). In results not reported, I find that controlling for an index of inexpensive quality measures does not change the corruption results presented above. This suggests that the results are actually being driven a reduction in corruption per se, rather than an overall change in the competence of those building the project.

6.2.2. Nepotism

The BPKP auditors examined the project’s financial records and inspected the construction site. They did not, however, examine who worked on the project, and whether those who worked had family ties to the officials running the project. To the extent that giving jobs to

\textsuperscript{24} This difference is statistically significant – a t-test in control villages rejects the equality of corruption in prices and quantities with a p-value of less than 0.01.
family members is an alternative method of extracting rents from the project, we might expect this type of non-audited behavior to respond to the increased audits. As discussed in Section 3, the sign of this response would depend on whether nepotism and direct theft of funds are substitutes or complements. Of course, there are other reasons to hire family members besides rent-extraction; for example, we might also expect work on the project by family members to increase if project officials wanted to improve the project and if family members were less prone to moral hazard than non family members.

I examine the change in employment by family members using data from the household survey described in more detail in Appendix A. Each respondent in the household survey was asked if he or she was related to any of seven types of village government members or the head of the project implementation team. I examine whether people who said that they were either immediate or extended family members of village government were more or less likely to report having worked for pay on the road project in audited villages than in control villages. Specifically, I estimate the following linear probability model using OLS:

\[
\text{WORKED}_{hijk} = \gamma_1 + \gamma_2 \text{AUDIT}_{jk} + \gamma_3 \text{FAMILY}_{hijk} + \gamma_4 \text{AUDIT} \times \text{FAMILY}_{ijk} + \gamma_5 \text{X}_{hijk} + \epsilon_{hijk}
\] (3)

where FAMILY is a dummy equal to 1 if the individual was a family member of village government or the head of the project, WORKED is a dummy equal to 1 if a household member worked for pay on the project, X is a vector of control variables (age and gender of respondent, predicted household income, and dummies for the ways the household was sampled), h represents the household, i the village, j the subdistrict, k represents the audit stratum, and \(\gamma_k\) is a stratum fixed effect. The coefficient of interest is \(\gamma_4\), which represents the differential probability

\[25\] Regressions using a Probit specification produce essentially similar results.
in audited villages relative to control villages that family members of village government or the head of the project worked on the project. The empirical results include two different FAMILY variables, one for being a family member of a government official and one for being a family member of the project head.

The results are reported in Table 4. The estimates suggest that family members of the village government were 7-8 percentage point more likely to work on the project in audited villages than in non-audited villages, and family members of the project head were 11-14 percentage points more likely to work on the project in audited villages than in non-audited villages. Given that the mean probability of working on the project is only 25 percent, these effects are quite large in magnitude, and suggest a substantial increase in nepotism in response to the audits.26

The two different interpretations for the change in employment by family members have very different implications – one view says that this is alternative form of corruption, whereas the other suggests this is actually an attempt to improve the project. Though distinguishing between these alternative hypotheses is difficult, there is some suggestive evidence in favor of the nepotism-as-corruption view. In particular, the micro-finance literature has suggested that social connections can be an effective mechanism for minimizing moral hazard (Karlan 2004), so if reducing moral hazard was the issue, one might expect similar effects for workers with many social connections as for family members. In column (3) of Table 4, however, I find that there is no differential effect in response to the audits in the relationship between social connections and working on the project. Furthermore, in results not reported, I find that, conditional on observables, family members of village officials are more likely to be employed in the higher

26 In results not reported, I find that there were no statistically significant changes in nepotism associated with the invitations and comment form treatments, although the point estimates suggest that there may have been some increase in nepotism in response to the comment forms.
wage category (skilled labor rather than unskilled), suggesting that they may be receiving rents from the project. While this evidence is suggestive of a nepotism-as-corruption story, it is by no means definitive, and understanding this phenomenon is an important direction for future work.

6.3. The Participation Experiments

6.3.1. Did the interventions increase participation?

Before we can assess the impact of increased participation on corruption, it is important to make sure that the treatments — i.e., the invitations and comment forms — did, in fact, increase villager participation. Table 5 examines the effects of the invitations and comment forms on several measures of participation in the accountability process. The data come from the meeting survey, so that each observation represents one village meeting. The sample includes all accountability meetings, so there are three observations for each village.

Column (1) examines the impact of the treatments on meeting attendance. The results show that the treatments had a substantial effect — the invitations treatment increased attendance at the meetings by an average of 13.4 people, or approximately 35 percent. Interestingly, it appears that being able to submit written comments and attending meetings are substitutes rather than complements. In particular, attendance at meetings was lower in villages that received both invitations and comment forms than in villages receiving invitations alone.

Second, using data from the attendance list, I classify people as “non-elite” if they have no official position in the village, no official position on the project, and were not described as a tokoh masyarakat (“informal village leader”) by village members who assisted the enumerator. Using this classification, the results in Column (2) show that virtually all of the increase in attendance at the meetings came in the form of increased attendance by these non-elite villagers, so that the number of non-elite at the meetings increased by over 75 percent. Taken together, the results suggest that the invitations treatment increased the percentage of non-elite villagers, who
may be less beholden to the village head than members of the village elite, from a minority of attendees at the accountability meetings to a majority.

Columns (3) and (4) examine the impact of the treatments on active participation at the meetings. Column (3) shows that, in the invitations treatment, the average number of people who spoke at a meeting increased by 0.60, or just under 10 percent. Column (4) shows that about half of these new speakers were non-elite, increasing the number of non-elite villagers who spoke at a meeting by almost 50 percent.

In addition, the comment forms appear to have been quite successful in eliciting villagers’ opinions about the project. On average, 140 comment forms (about 35 percent) were returned, filled out, per meeting. Of these, approximately 60 percent of the forms included not only answers to the closed-response questions, but also included handwritten comments in response to one or both of the open-ended questions. The responses were quite varied, and on average had slightly more positive than negative comments. There are no substantial differences in the response rate across villages of differing average education levels—mean adult education in these villages is 4.8 years, and 89 percent of adults can read and write—which suggests that, at least within the level of education in rural Java, education does not seem to be a substantial constraint to using comment forms to elicit villager responses.

6.3.2. Effect on Meetings

Table 6 investigates the effect that increased participation (via the invitations and comment forms) had on the accountability meetings. As discussed in more detail in Appendix A, the enumerator recorded each problem that was discussed at the meeting, where a “problem” was defined as the topic of any substantial discussion other than the routine business of the meeting. The enumerator also assigned one of 57 problem codes to each problem, and coded whether the
problem was corruption-related or not. In addition, I define a ‘serious response’ to a problem as any of the serious actions that could be taken by a village in response to a problem with the project – specifically, agreeing to replace a supplier or village official, agreeing that money should be returned, agreeing for an internal village investigation, asking for help from district project officials, or requesting an external audit. These serious actions are quite rare – they occur at only 2 percent of meetings – and thus, to preserve statistical power, I consider them together.

The results in Column (1) suggest that neither the invitations treatment nor the invitations + comment form treatment had a significant effect on the total number of problems discussed at the meeting – the point estimates suggest an increase of .07 to .11 problems per meeting, or less than 10%, and are not statistically significant. However, as shown in Column (2), the invitations appear to have increased the probability of having a corruption-related problem discussed at the meeting by about 3 percentage points, or 50 percent above the baseline level.

Interestingly, relative to having invitations only, adding the comment forms did not change the probability of a corruption-related problem being discussed at the meeting. However, the comment forms do make a difference in how these problems are resolved. Columns (3) and (4) show that the probability of a serious action being taken to resolve a problem is 2 percentage points higher – or double the baseline amount – in villages receiving the comment forms than in those that did not. Among the (albeit endogenously determined) set of villages with a corruption-related problem, villages in the comment form treatment were 11 percentage points more likely to have a serious response than those with either invitations alone or with neither invitations nor

---

27 Classifying problems and listing whether they are potentially corruption related clearly requires some degree of subjective judgment on the part of the enumerator filling out the form. However, all villages in a subdistrict were handled by the same enumerator, so including stratum (i.e., subdistrict) fixed effects controls for these potential differences in coding.
comment forms.\textsuperscript{28} Though these effects are statistically significant, it is important to note that, in absolute terms, they are small in magnitude.

These results suggest that the invitations and comment forms had very different, and to some degree, unexpected effects on the meetings. Increasing the number of outsiders at the meeting through the invitations treatment appears to have increased the probability of a corruption-related problem being discussed at the meeting. Adding the comment forms, on the other hand, changed not the types of problems being discussed, but rather the probability that, conditional on a problem arising, a serious action was taken to resolve it. This suggests that the main effect of the comment forms may be more about creating common knowledge about the problem, rather than affecting first-order knowledge about the problems per se. Put another way, when an issue was brought up, possibly by an elite member of the village who had less fear of retaliation by the village head, villagers may have been unsure whether to side with the challenger or the village head. Knowing from the comment forms that many other people agreed with them, and therefore that the challenge was likely to be victorious, may have tipped the balance.

6.3.3. Main Theft Results

Table 7 examines the overall impact of the two participation treatments on corruption in the projects. The first panel shows the effect of the invitations treatment; the second panel shows the effect of the invitations + comment forms treatment. The results show that neither the invitations nor the invitations + comment forms treatments had a substantial impact on overall

\textsuperscript{28} One potential concern about these results is that they may be driven by differences in the problems that arise. To investigate this, in results not presented I repeat the analysis on the problem level, and include fixed effects for each of the 57 possible problem codes. The results are very similar to the meeting-level results presented in Table 6, which suggests that, to a reasonably fine level, the results are not being caused by changes in the composition of problems. Also, note that columns (5) and (6) do not include subdistrict fixed effects. This is because it is rare to have more than 1 village in a subdistrict with a corruption-related problem, so including subdistrict fixed effects reduces the effective sample size dramatically.
corruption. Depending on the specification and the measure of corruption, the point estimates for the two treatments range from an increase in corruption of 0.8 percentage points to a reduction in corruption of 3.8 percentage points, and are never statistically distinguishable from 0.

The lack of a strong effect of the invitations and comment forms on corruption is consistent with the evidence in Section 6.3.2, for while the comment forms did increase the probability of serious action being taken in response to corruption, the magnitude of this response was small. In fact, even if taking a serious anti-corruption action at a meeting eliminated corruption entirely, the reduction in corruption in caused by these induces serious actions would have been less than 1 percentage point of average total expenditures.\(^29\) The results here suggest that not only were the direct effects (via anti-corruption actions) small, but any deterrent effects of the treatments were small as well.

Of course, even without the invitations and comment forms, the KDP program investigated here includes more grass-roots participation, and a more complicated system of checks and balances, than the typical government project in most developing countries. It is possible that the marginal effects of increasing participation from this relatively high baseline—i.e., the effects reported above—do not capture the overall effects of grass-roots participation as a monitoring mechanism, and that more dramatic variations in the amount of participation might have different effects.

However, even though increasing participation did not change the overall level of corruption, it did affect the type of corruption chosen by village officials, which suggests that the

\(^{29}\) To see this, suppose that taking a serious anti-corruption action eliminates corruption in the village. (In fact, in the cross-section, villages where a serious response was taken had corruption levels about 23 percentage points lower than other villages.) If I re-estimate Table 6 at the village level, rather than the meeting level, the point-estimate is that comment forms treatment increased the probability of a serious response to a problem by 3.3 percentage points. This suggests that, if the sole effect of the comment forms was through the probability of a serious response being taken, the expected average reduction in corruption would be \(23 \times 0.033 = 0.75\) percentage points.
changes in participation induced by the interventions were large enough to make a difference. As shown in Table 7, the invitations treatment led to a statistically significant reduction in corruption on the labor dimension of between 23 percentage and 26 percentage points. The invitations + comment forms also led to a reduction in corruption in labor, though it was somewhat smaller in magnitude (10 to 13 percentage points) and not statistically significant.

These reductions in corruption on the labor dimension appear to have been largely offset by increases in corruption in materials such as rocks and gravel. Specifically, the point estimates suggest a 6.4 percentage point increase in materials associated with the invitations treatment. Since, as shown in Table 2, these materials account for 68% of road expenditures (as compared to only 20% for unskilled labor), this 6.4 increase in theft of materials is large enough to offset the 23 percentage point decline in theft of wages, and to generate the 0 total change in overall theft.30

An important question is the whether this shift in corruption methods represent a change in actual expenditures or simply reflect changes in accounting by the villages. To examine this, for both materials and unskilled labor, I compute the difference between the log of the estimated actual expenditures (as estimated from the engineering, price, and worker surveys) and the log of the planned expenditures. The planned expenditures are from the initial proposed budget for the project, which was fixed before the randomization was announced, and is therefore exogenous with respect to the treatments. I compute an analogous measure for the final reported expenditures (i.e., the difference between the log of reported expenditures for materials, unskilled labor, and other expenses and the log of the corresponding amounts from the planned

30 As discussed above, the change in log (materials + wages) and log(materials) + log(wages) will not be identical. This, combined with treatment effect heterogeneity, may explain why for the invitations + comments treatments, the point estimates for the change in theft of materials is actually small but negative, whereas mechanically it must be positive to offset the decline in theft of wages.
budget.) To ensure that the results capture substitution only, I restrict the sample to the 406 villages for which all variables are non-missing, so that the sample is the same for all 5 variables.

Table 8 reports the changes in actual vs. reported expenditures for the invitations treatment. The results indicate that the substitution from theft of wages to theft of materials identified above was predominantly an accounting change in where in the books the corruption was hidden, rather than a change in real expenditures. Specifically, relative to the planned amounts, there was no change in the actual amount of materials purchased, and only a slight (but not statistically significant) increase in the amount actually spent on wages. By contrast, there were much sharper, and statistically significant, changes in how expenditures were reported. In particular, the results indicate increases of between 3.5 and 5.9 percentage points in the amount of reported expenditures on materials and between a 1.1 and 7.1 percentage point increase in the amount of reported ‘other’ (i.e., other than rocks, sand, gravel, and unskilled labor) expenditures; by contrast, there was an offsetting decrease in the reported amount of wage payments of between 11.9 and 17.5 percentage points. Although not reported, the invitations + comment forms treatment shows qualitatively similar results, but smaller in magnitude and not statistically significant.31

These results suggest that the shift in corruption methods was nominal, not real. More generally, this suggests that the technology of corruption is such that the decisions of how much to actually spend on a given line-item and how much corruption to mask in that line-item may be, broadly speaking, separable decisions. Thus, a focused effort to reduce corruption in a particular area, even if successful, will not necessarily increase real expenditures in that area. The implications of this will be discussed in more detail in the cost-benefit analysis below.

---

31 One can also do a similar exercise for the audits treatment. In results not reported, I find that the decrease in corruption due to the audits was entirely due to an increase in actual expenditures on both materials and wages.
There are several different explanations for the observed substitution from corruption in labor to corruption in materials. First, this substitution may have occurred because corruption in labor was technologically easier for villagers to detect, as it may be easier for villagers to observe wage payments than the quantity of materials delivered. Alternatively, even if villagers had equal information about both types of corruption, the substitution may have occurred because the invitations induced more workers to attend the meetings, and those workers focused on their private interest (i.e., the wages they personally were supposed to be paid by the project) rather than the public good of a higher quality road.

Though these hypotheses are difficult to definitively distinguish, there is suggestive evidence in favor of the latter hypothesis. To see this, one can examine whether the invitations treatment varies when workers come from outside the village, and are therefore ineligible to attend the accountability meetings. This relationship is shown in Figure 1, which presents the results from a Fan (1992) locally-weighted regression of the theft of wages on the percentage of workers from outside the village. Figure 1 shows that effect of the invitations on reducing theft of wages, i.e. the difference between the theft of wages in invitations and control villages, declines as the percentage of workers from outside the village increases. On the other hand, in results not reported, when I examine household’s knowledge about wages and corruption in the project, I find that the quality of this information does not change as the percentage of workers from outside the village increases.

Since the only people that attend the accountability meetings are those who live in the village, these results suggest that it is the fact that the invitations induce workers to attend

---

32 The Fan locally-weighted regression, a non-parametric estimation technique, estimates a separate regression at each point on the x-axis, where for each regression, observations close to that point are weighted more than observations further away. More details can be found in Deaton (1997). In this case, separate Fan regressions were run for invitations and controls. Figure 1 plots the predicted values from each of these regressions.
meetings and fight for their own wages that is driving the substitution away from theft in wages. Even if information played a role as well, the results suggest that grass-roots monitoring is likely to be most effective in reducing corruption in cases where villagers have both direct information and strong personal incentives to monitor officials, and less effective when they are asked to monitor technically complex public goods.

7. **Cost-Benefit Analysis**

The previous analysis discussed the direct effects of the treatments. It showed that the audit treatment reduced corruption, and that the invitations treatment changed the method through which corruption was hidden. This section performs a cost-benefit calculation to assess whether, on net, the benefits from these treatments exceed the costs. This exercise requires making several assumptions, particularly about the efficiency cost of different types of corruption, and therefore should be viewed as somewhat more speculative than the preceding sections.

Table 9 presents the cost-benefit estimates for the audits and invitations treatments. (Since none of the estimated effects of invitations + comment forms were statistically significant, I omit them from this analysis.) For each treatment, I present two sets of net benefits – “equally weighted net benefits,” calculated under the assumption that the marginal utility of income is constant across individuals, and “distribution weighted net benefits,” which takes into account the fact that some benefits and costs are borne by the rich while others are borne by the poor. The net benefits are derived as follows. First, the monetary cost of the audits is the actual cost paid by the project per audit, including the salary of the auditors, and the monetary cost of the invitations is the actual cost paid for photocopying the invitations. The associated dead-

---

33 Specifically, the distribution-weighted net benefits assumes CRRA utility of per-capita consumption with a coefficient of relative-risk aversion of 2, normalized so that the median household in rural Java has marginal utility of 1.
weight loss is the dead-weight loss associated with the increased taxes required to pay the monetary cost of the treatments. In line with various estimates, I assume that the marginal cost of public funds is 1.4 (i.e. the dead-weight loss is 0.4). Finally, the time cost for both treatments is the monetary value of the additional time villagers spend at village meetings as a result of the treatments, valued at average local wage rates.

The estimates for the change in rents received by corrupt officials are taken from the point estimates in Table 3 and Table 7, multiplied by the average cost of the project and the average percent of the project consisting of materials and labor expenditures from Table 2.

As discussed above, a change in a particular type of corruption does not necessarily imply a change in real expenditures in that area. Therefore, to estimate the change in benefits from the project, I use the point estimates of the change in actual expenditures. For wages, the increase in actual wage expenditures is valued at the actual amount transferred. For materials, the increase in materials expenditure increases the lifespan of the road. To value the change in lifespan of the road, I use estimates from a cost-benefit evaluation of KDP roads conducted for the World Bank (Dent 2001), which imply that the marginal dollar of materials stolen reduces the discounted benefits from the road by 3.41 dollars.

To include distributional concerns, I make the following assumptions. First, I estimate that households where at least one household member is in the village government have per-

---

34 Estimates of the marginal cost of public funds for indirect taxes vary considerably depending on the country and the method used. For developing countries, estimates for the marginal cost of funds from indirect taxes range from a low estimate of 1.04 to 1.05 in Indonesia and Bangladesh (Devarajan et. al, 1999) to between 1.59 and 2.15 for India (Ahmad and Stern, 1987). By comparison, estimates of the marginal cost of public funds for the U.S. range from 1.17 to 1.56 (Ballard et. al, 1985), with policy analysis typically using values in the range from 1.30 to 1.40.

35 Dent estimates that KDP road projects produces annual flow benefits of 33% of their cost each year for the life of the road. He also estimates that a good quality road will last 10 years and a poor quality road will last 5 years. Based on conversations with KDP engineers, I assume that each percentage of materials stolen reduces the life of the road by 0.1 years, so stealing 50% of materials is the difference between ‘good’ and ‘poor’ quality. Assuming a real discount rate of 5% and that the baseline lifespan of a road is 7 years implies that each dollar of materials stolen reduces the discounted net benefits of the project by 3.41 dollars.
capita expenditure 18.5 percent higher than typical households in the village. Assuming the consumption levels of project officials is similar to village officials, this implies (given CRRA utility with a coefficient of relative risk aversion of 2) that the social value of $1 of these rents is $0.61. On the other hand, the social value of $1 of increased wages received by workers is $1.29, as workers have per-capita consumption approximately 13 percent lower than the median in the village. I assume that the social benefits from the road are enjoyed equally by all in the village, so they have a marginal social value of 1. Finally, for taxes, I estimate the distributional impact of both the monetary and deadweight loss cost using national consumption data from the 2003 SUSENAS, assuming that the burden is borne proportionally to consumption.

The results presented in Table 9 suggest that the audits were substantially cost-effective. I estimate that the net social benefits from the audits were approximately $655 per village, which implies that the benefits were more than double the costs of the audits. Weighting by distributional incidence increases these net benefits even more.

Interestingly, I also estimate a positive net social benefit associated with the invitations. However, this is largely driven by the very small estimated increase in actual expenditures of materials reported in Table 8 – a very statistically insignificant 1.4 percentage points. If I used 0 instead of 1.4 for this point estimate, then the estimated net benefits from the invitation becomes very close to 0.

The audit treatment discussed here was a move from a 4% baseline audit probability to a 100% audit probability. It is possible, however, that the response of corruption to the audit probability is concave, so that an audit probability of only 50% or even 25% might achieve most of the benefits.\textsuperscript{36} The costs of audits, however, are roughly linear in the audit probability. This

\textsuperscript{36} For example, Nagin. et al. (2002), in a study of monitoring of call center employees, found substantial evidence of diminishing returns to increasing the monitoring probability.
suggests that raising audit probabilities to an intermediate level, rather than all the way to 100%,
might have an even higher cost-benefit ratio than the 100% audit probability documented here.

8. Conclusion

This paper has examined the results of a field experiment in Indonesia, designed to
investigate alternative approaches to fighting corruption. I examined the effect of two strategies:
top-down monitoring by government auditors, and bottom-up monitoring through grass-roots
participation in the village monitoring process. I find considerable evidence that increasing the
probability of external audits substantially reduced the amount of theft in the project. A cost
benefit analysis suggests that the benefits from the audits exceeded their cost.

By contrast, increasing grass-roots participation in monitoring the project altered the
method of corruption, but had minimal effects on the overall level of corruption. In particular,
the main response to increasing grass-roots participation seems to have been a shift in where
corruption was hidden – away from hiding corruption in wages and towards hiding corruption in
over-invoiced materials procured for the project. These results suggest that grass-roots
monitoring may be more effective for government programs that provide private goods, such as
subsidized food, education or medical care, where individual citizens have a personal stake in
ensuring that the goods are delivered and that theft is minimized. For public goods, such as the
infrastructure projects studied here, the results suggest that using professional auditors may be
more effective.

This paper suggests several interesting areas for future work. First, all of the analysis in
this paper was based on measurements of actual corruption. While this study was able to use
direct measures of corruption, doing so is always difficult and, in many other settings, may not
be possible. In those cases, as in much of the existing research on corruption, researchers may be
forced to use data on people’s perceptions of corruption rather than direct measures of corruption. Little is known about the relationship between these perceptions and reality, and potential biases may be severe.\textsuperscript{37} For example, well-publicized punishments of corrupt officials may reduce subsequent corruption but increase people’s awareness of corruption. In that case, perceptions of corruption and actual corruption could actually be inversely correlated.

In order to explore this issue, the household survey conducted as part of this study included questions in which villagers were asked their perceptions of corruption, both in Indonesia in general and in the road project in their village specifically. I can match this data with the actual data on corruption to examine the relationship between perceived corruption and reality. I consider this a promising direction for future work.

Another interesting area for future work is to explore the long-run effects of attempts to combat corruption. For example, if auditors are bribable, over time villages may develop repeat relationships with auditors which may make bribing auditors easier than in the one-shot case examined here. This might suggest, for example, that frequent rotation of auditors – or lower probabilities of audits combined with higher punishments – may be optimal. Understanding the long-run implications of anti-corruption policies, and developing ways of ensuring that they remain effective over the long term, is an important issue that remains to be explored.

\textsuperscript{37} The one exception is a recent paper by Mocan (2004), who examined the relationship, at the cross-country level, between perceptions of corruption in a country and micro data on the frequency that people in that country paid bribes.
References


Bardhan, Pranab and Dilip Mookherjee, “Decentralization and Accountability in Infrastructure Delivery in Developing Countries,” mimeo, Boston University, 2003.


Appendix A  Data collection

In addition to the corruption field survey described in Section 5 above, this paper uses three other types of data collected during the course of the project—key-informant surveys, data on village meetings, and a household survey. This Appendix describes the data on village meetings and the household survey. In addition, it provides additional details on the corruption field survey described above.

Village meeting data

In each KDP year in a village, there are a total of seven regularly scheduled village meetings that must occur (including the three accountability meetings.) The enumerator was present at the final four meetings – the meeting at which preparations for construction were began plus the three accountability meetings. At each meeting, the enumerator circulated an attendance list, which asked each attendee to list his or her name, gender, age, hamlet of residence, and position in the village (if he or she had one). An assistant was present to assist villagers filling out this form. As the meeting progressed, the enumerator (with the assistance of a local counterpart) noted on the attendance list each person who spoke at the meeting. In addition, sitting with the local counterpart after the meeting was over, the enumerator asked the local counterpart to identify which of the attendees was a tokoh masyarakat, or informal leader, a reasonably objective categorization typically given to teachers, religious leaders, or other types of village elders.

While the meeting was in progress, the enumerator was asked to keep detailed notes on what occurred during the meeting. Afterwards, the enumerator was asked to compile two lists. The first list was a list of each item on the official meeting agenda. For each agenda item, the enumerator was asked to list the time the agenda item began to be discussed and the time it finished, who spoke during the discussion of that agenda item, and whether there were any problems that arose or decisions that were taken as part of that agenda item. The second list was a list of all problems that arose at the meeting. A “problem” was defined as the topic of any substantial discussion other than the routine business of the meeting; the median problem reported in the data was discussed for 7 minutes, and the mean number of problems reported in an Accountability Meeting was 0.73. For each problem, the enumerator described the problem, classified it according to one of 57 pre-defined problem codes, and listed the amount of time spent discussing the problem, who first raised the problem, who was potentially involved in the problem, whether there were indications of corruption in the problem, whether the problem was resolved, and if so, what actions were taken to resolve it.

Household survey

The household survey was conducted approximately during the last two months of construction and the first month after construction was completed. The survey contained a household roster, a list of assets, information on participation in social, religious, and government activities, detailed information on participation in the road project, and a series of questions about perceptions of corruption.

To estimate household expenditure of respondents, I used the 1999 SSD (Hundred Villages Survey), an Indonesian statistics bureau dataset, containing 3,193 rural Javanese households. The SSD asked both a detailed expenditure questionnaire and the same set of asset
questions used in my household survey. In the SSD, I used OLS to estimate the relationship between log household expenditure and the following variables, all of which I observe in my survey: log household size, whether the household was headed by a woman, the percentage of household members consisting of children ages 0-3, 4-6, 7-9, 10-12, and 13-16, dummies for whether the household has a stove, refrigerator, radio, television, satellite dish, motorbike, car, and electricity, dummies for floor type, wall type, and ceiling type, the total amount of land held by the household, whether the household consumes meat at least once a week, whether each household member has at least two sets of clothes, whether the household uses modern medicine when a child is sick. I then used the estimated coefficients from the SSD to predict household expenditure in my survey. Combined, these 34 variables have an R-squared of 0.58 predicting log household expenditure in the SSD, which suggests that predicted expenditure is a reasonable approximation for actual expenditure, at least for the purposes used here.

The household survey was designed as a stratified random sample, containing between six and thirteen respondents per village, selected as follows. Two respondents were drawn randomly from the attendance list of Village Meeting II, which was held before the randomization was announced, and is therefore exogenous with respect to the experiments. Two respondents were selected from the hamlets in which the road was located by first randomly selecting a hamlet, and then randomly selecting a neighborhood (RT) in that hamlet. A complete list of households in the RT was obtained from the neighborhood head, and two households were drawn randomly from that list. Individual respondents were drawn from the a list of all adults age 18 or over in the selected households. Two additional respondents were selected from the hamlets in which the road was not being built using the same procedure. As men in the village tend to participate much more in road construction activities, the randomization was designed such that, of the four respondents selected in this manner, three were men and one was a women. In villages receiving the Comment Form treatment, an additional four respondents were drawn using the same procedure, two from hamlets with the project and two from hamlets that did not contain the project. Each respondent received compensation of Rp. 10,000 ($1.20), equal to slightly more than half of the typical daily agricultural wage in the study area.

An additional goal of the project was to measure how corruption perceptions change when individuals are told that the perceptions will actually be used as part of monitoring, compared to when they are responding for survey purposes only. To examine this, at the very end of the survey (after all questions except for questions involving corruption in the project and the lottery question had been asked), the additional 4 respondents in the Comment Form villages were told that their responses to the questions about corruption in the project would be used, anonymously, as part of the overall report on the comment forms presented at the accountability meeting. To simplify exposition, I will refer to this variant of the questionnaire as Form B, and to the normal questionnaire, in which all questions were explained to be anonymous and used for research purposes only, as Form A. The design was therefore that all villages would have six households receiving Form A surveys, and that in each of the Comment Form villages, there would be an additional 4 households receiving Form B surveys.

Due to a training error, approximately 60% of enumerators appear to have given Form B surveys to all households in Comment Form villages. Therefore, in approximately half of all Comment Form villages, three additional households were surveyed, drawn randomly from the same neighborhoods as before. Of these three additional households, two were randomly assigned to receive Form A and one was assigned to receive Form B. In analysis using these corruption questions, I therefore include controls for whether the household received Form B and
for whether the household was sampled as part of this additional three households per village sampled. Under the assumption that this enumerator error was orthogonal to the error term, this will allow me to separately identify the effect of being in a Comment Form village, the effect of receiving a Form B form, and the effect of being in the late sample.

Field survey / Corruption measurement

The general approach used in the field survey is described in Section 5 of the text above. This section discusses a number of additional aspects of the data collection not discussed above that are important for the analysis in this paper.

One important issue is the treatment of voluntary contributions. According to official village reports, these contributions account for an average of 16% of total project costs. Of these voluntary contributions, the bulk (60% according to official village reports) comes in the form of either voluntary labor or tools (typically, village workers bring their own tools to work on the project; the value of these tools is often reported as a voluntary contribution). Anecdotal evidence confirms that voluntary contributions are qualitatively important primarily for labor and tools.

I treat these voluntary contributions as follows. For the three main materials in the road projects—sand, rocks, and gravel—I include both reported project expenditures and reported voluntary contributions when calculating the total amount of reported materials, as this total amount should be comparable to the total amount of each type of material observed in the field. For labor, where overstating voluntary labor is a much more important potential margin of corruption than in materials, a different approach was taken. As discussed in Section 5, for labor, I use only reports and actual estimates of paid labor, and exclude voluntary labor.

For the worker survey described in Section 5 above, a total of five worker interviews were conducted in each village. Of these interviews, two were focus group interviews, consisting of 3 or more workers interviewed together, and three were individual interviews. Workers were selected to be interviewed as follows. Ten workers were randomly drawn from the official list of all workers who had ever been paid by the project. Surveyors then went to find, in order, these workers, until two were found for interviews. For the first worker found, the worker was asked to gather at least two other people who had worked on the project with him for a focus group interview; the second worker found was interviewed alone. Since the worker lists may not be complete (and, in particular, may be a non-random sample of workers), for the remaining three interviews (one focus group and two individual), workers were recruited more informally, by having the surveyor go to different areas of the village located near the road and asking households who had worked on the project.

The responses on the worker survey are in some cases quite variable. In particular, there is often variation because certain workers only worked on certain aspects of the project, because certain parts of the project were done differently than others, or because some workers were confused about the meaning of the questions. Because these discrepancies can be quite difficult to capture on the survey, surveyors were asked to fill out a form in which they summarized the results of all interviews. The surveyors were instructed that the summary should represent their best understanding of what actually happened in the village, based only on the information in the worker surveys. Though this summary is necessarily more subjective than answers to the worker surveys, experience during the pilots suggests that this method is more accurate than using mechanical averages from the individual worker surveys. Accordingly, the main results use the information from this summary report.
Appendix B  Assumptions and Calibration

Two main types of assumptions are used in the corruption calculations in this paper—assumptions about loss ratios, defined as the percentage of materials unaccounted for as a result of normal construction losses and measurement error, and assumptions about worker capacity, defined as the amount of each type of task an average worker can accomplish in a day of work. This section describes the assumptions used in the paper in more details, and discusses the calibration exercises through which these assumptions were determined.

To determine the appropriate assumptions, two methods were used. First, similar assumptions are used in the planning process for KDP roads. For example, KDP engineers typically assume normal construction losses of 16 percent for sand and gravel, a 23 percent loss of rock. Therefore, when purchasing supplies, engineers increase the final volume of materials they need by these amounts to ensure the appropriate quantities at the time of construction. Similarly, when budgeting labor requirements, KDP engineers estimate the quantities of each type of task in the project, and multiply by standard worker capacity estimates. As a baseline, I therefore obtained, from the KDP chief engineer, the standard assumptions typically used for KDP projects.

There are several problems with these assumptions, however. For the loss ratios for physical materials, the estimates included only the losses during the construction process; they do not include the additional losses inherent in the measurement process. They are therefore likely to be too small for the purposes here. For manual labor, the estimates used were apparently based on a 1970s workfare program, when both worker nourishment and worker motivation were substantially lower than today.

Given this, I conducted 4 calibration exercises. In each calibration exercise, our project built a 60 meter Telford road, similar in construction techniques and standards to KDP road projects. Materials were purchased from local suppliers, and labor was recruited from the villages in which the roads were constructed. The roads were constructed in four very different regions of the study area, which were chosen to represent the different types of conditions typical in the study area. Detailed measurements were taken of all materials delivered to the site, and careful track was kept of each worker’s activities during construction.

Once the road was completed, the measurement techniques from the study were applied to the test road. Using these techniques, I estimated the total quantity of material used in each road. By comparing the quantity of material estimated using the study techniques to the actual quantity of material we used in the road (which is known, since it was measured as it arrived at the project site), one can recover the correct loss ratios.

From the logs of worker activity, I was able to construct the actual time required by the workers on these test roads to complete each of the tasks assigned to them. Of course, workers who know that they are being closely watched may work more quickly than normal workers in the field. Nevertheless, the pace of work in the test road was between 50 and 700 percent faster, depending on the task, than the standard assumptions. Additional conversations with field engineers confirmed that the 1970s standards were, in fact, quite low. As one engineer said, the standards were so loose that “if a project using those standards didn’t finish with money left over, I was immediately suspicious that there was corruption in that village.”

Table 10 lists the main assumptions used in the study. For each assumption, it lists the original assumptions from project engineers, the results of the calibration exercises, and the revised assumptions used in the paper. The revised assumptions were arrived at after extensive discussions with field engineers, incorporating the results of the test survey. For the labor
estimates, the revised estimates were revised downward 20% from the estimates from the test road, to incorporate breaks taken by workers and to take into account the fact that carefully watched workers may work faster than workers under more normal monitoring conditions.

Interestingly, the high loss ratio for gravel and low loss ratio for sand suggests that some of the gravel is seeping through cracks in the rocks, and is counted by our survey as sand. This suggests that one might better consider “sand + gravel” together (with an implied joint loss ratio of 1.25), rather than separately. Doing so reduces the average percent missing from 25 percent to 23 percent, but otherwise does not substantially alter the results of the paper.
### Table 1: Number of villages in each treatment category

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>Invitations</th>
<th>Invitations + Comment Forms</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>114</td>
<td>105</td>
<td>106</td>
<td>325</td>
</tr>
<tr>
<td>Audit</td>
<td>93</td>
<td>94</td>
<td>96</td>
<td>283</td>
</tr>
<tr>
<td>Total</td>
<td>207</td>
<td>199</td>
<td>202</td>
<td>608</td>
</tr>
</tbody>
</table>

Notes: Tabulations from results of randomization. Each subdistrict faced a 48% chance of being randomized into the Audit treatment. Each village faced a 33% chance of being randomized into the Invitations treatment, and a 33% chance of being randomized into the Invitations + Comment Forms treatment. The randomization into Audits was independent of the randomization into Invitations or Invitations + Comment Forms.

### Table 2: Summary Statistics

<table>
<thead>
<tr>
<th></th>
<th>Total project size (US$)</th>
<th>Share of total reported expenses</th>
<th>Share of reported road expenses</th>
<th>Number observations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>8,858</td>
<td>0.764 (0.231)</td>
<td>0.094 (0.074)</td>
<td>556</td>
</tr>
<tr>
<td></td>
<td>(4,388)</td>
<td>0.154 (0.180)</td>
<td>0.473 (0.161)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.082 (0.170)</td>
<td>0.117 (0.187)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.116 (0.188)</td>
<td>0.116 (0.148)</td>
<td></td>
</tr>
</tbody>
</table>

### Table 3: Audits – main theft results

<table>
<thead>
<tr>
<th>Percent missing:</th>
<th>Control Mean</th>
<th>Treatment Mean: Audits</th>
<th>No Fixed Effects</th>
<th>Engineer Fixed Effects</th>
<th>Stratum Fixed Effects</th>
<th>Num Obs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log reported value – Log actual value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major items in roads</td>
<td>0.296 (0.036)</td>
<td>0.211 (0.033)</td>
<td>-0.086 (0.049)</td>
<td>0.083 (0.044)</td>
<td>-0.090 (0.044)</td>
<td>-0.056 (0.038)</td>
</tr>
<tr>
<td>Major items in roads and ancillary projects</td>
<td>0.296 (0.031)</td>
<td>0.218 (0.034)</td>
<td>-0.077 (0.046)</td>
<td>0.098 (0.042)</td>
<td>-0.079 (0.042)</td>
<td>0.061 (0.037)</td>
</tr>
<tr>
<td>Breakdown of roads:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Materials</td>
<td>0.223 (0.043)</td>
<td>0.175 (0.040)</td>
<td>-0.049 (0.058)</td>
<td>0.404 (0.054)</td>
<td>-0.031 (0.054)</td>
<td>0.561 (0.050)</td>
</tr>
<tr>
<td>Unskilled labor</td>
<td>0.333 (0.086)</td>
<td>0.263 (0.079)</td>
<td>-0.066 (0.117)</td>
<td>0.573 (0.096)</td>
<td>-0.102 (0.096)</td>
<td>0.293 (0.086)</td>
</tr>
</tbody>
</table>

Notes: Robust standard errors in parentheses. Audit effect and p-values are computed from a regression of the dependent variable on a dummy for audit treatment, invitations treatment and invitations + comment forms treatments, allowing for robust standard errors clustered by subdistrict to account for clustering of treatment by subdistrict. All dependent variables are the log of the value reported by the village less the log of the estimated actual value, which is approximately equal to the percent of corruption. Villages are included in each row only if there was positive reported expenditures for the dependent variable listed in that row.

### Table 4: Nepotism

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audit</td>
<td>-0.022</td>
<td>-0.005</td>
<td>-0.021</td>
<td>-0.043</td>
</tr>
<tr>
<td>Village Gov't Family Member</td>
<td>-0.026</td>
<td>0.013</td>
<td>0.013</td>
<td>-0.020</td>
</tr>
<tr>
<td>Project Head Family Member</td>
<td>0.035</td>
<td>-0.033</td>
<td>0.036</td>
<td>-0.020</td>
</tr>
<tr>
<td>Social activities</td>
<td>0.012***</td>
<td>0.012***</td>
<td>0.008</td>
<td>0.010*</td>
</tr>
<tr>
<td>Audit × Village Gov't Family Member</td>
<td>0.086***</td>
<td></td>
<td>0.071**</td>
<td></td>
</tr>
<tr>
<td>Audit × Project Head Family Member</td>
<td></td>
<td>0.142***</td>
<td>0.114**</td>
<td></td>
</tr>
<tr>
<td>Audit × Social activities</td>
<td></td>
<td></td>
<td>0.008</td>
<td>0.006</td>
</tr>
<tr>
<td>Stratum Fixed Effects</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Observations</td>
<td>4018</td>
<td>4018</td>
<td>4018</td>
<td>4018</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.21</td>
<td>0.21</td>
<td>0.21</td>
<td>0.21</td>
</tr>
<tr>
<td>Mean dep. variable</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
</tr>
</tbody>
</table>

Notes: Data is from the household survey. Each observation represents one household. Dependent variable is a dummy for whether a household member worked (for pay) on the road project. Estimation is by OLS with stratum fixed effects. Robust standard errors in parentheses, adjusted for clustering at subdistrict level. Social activities refers to the number of social activities adult household members participated in during the last month. All specifications include controls for invitations and invitations + comment form treatments, age and gender of respondent, mean adult education in the household, predicted household income, and dummies for type of household sampled.

* significant at 10%; ** significant at 5%; *** significant at 1%
Table 5: Participation – First stage

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attendance</td>
<td>13.359***</td>
<td>13.467***</td>
<td>0.602**</td>
<td>0.349***</td>
</tr>
<tr>
<td>Attendance of Non-Elite</td>
<td>(2.039)</td>
<td>(1.867)</td>
<td>(0.268)</td>
<td>(0.126)</td>
</tr>
<tr>
<td>Invitations + Comment</td>
<td>10.105***</td>
<td>10.214***</td>
<td>0.302</td>
<td>0.227**</td>
</tr>
<tr>
<td>Distribution by Village Gov’t</td>
<td>(2.296)</td>
<td>(2.171)</td>
<td>(0.229)</td>
<td>(0.107)</td>
</tr>
<tr>
<td>Distributed 500 invitations</td>
<td>0.740</td>
<td>-0.891</td>
<td>0.267</td>
<td>-0.041</td>
</tr>
<tr>
<td>Distribution by Village Gov’t</td>
<td>(1.998)</td>
<td>(1.855)</td>
<td>(0.216)</td>
<td>(0.107)</td>
</tr>
<tr>
<td>Meeting #2</td>
<td>-5.343***</td>
<td>-4.023***</td>
<td>0.202</td>
<td>0.065</td>
</tr>
<tr>
<td>Meeting #3</td>
<td>-4.486***</td>
<td>-5.953***</td>
<td>0.486***</td>
<td>-0.127</td>
</tr>
<tr>
<td>Observations</td>
<td>1769</td>
<td>1769</td>
<td>1769</td>
<td>1769</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.39</td>
<td>0.38</td>
<td>0.47</td>
<td>0.28</td>
</tr>
<tr>
<td>Mean dep. variable</td>
<td>47.88</td>
<td>24.10</td>
<td>7.99</td>
<td>0.95</td>
</tr>
<tr>
<td>Mean dep. variable</td>
<td>47.88</td>
<td>24.10</td>
<td>7.99</td>
<td>0.95</td>
</tr>
</tbody>
</table>

Notes: Data is from the meeting survey. Each observation is a single village meeting. Includes stratum (subdistrict) fixed effects. Robust standard errors in parentheses, adjusted for clustering at village level.

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 6: Participation – Impact on Meetings

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problems</td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
</tr>
<tr>
<td>Number problems</td>
<td>0.074</td>
<td>0.029**</td>
<td>-0.006</td>
<td>-0.000</td>
<td>0.008</td>
<td>0.010</td>
</tr>
<tr>
<td>Corruption-related problem</td>
<td>(0.067)</td>
<td>(0.014)</td>
<td>(0.007)</td>
<td>(0.006)</td>
<td>(0.048)</td>
<td>(0.048)</td>
</tr>
<tr>
<td>Serious response</td>
<td>0.107</td>
<td>0.026**</td>
<td>0.013*</td>
<td>0.014**</td>
<td>0.121*</td>
<td>0.101</td>
</tr>
<tr>
<td>Serious response (except audit)</td>
<td>(0.067)</td>
<td>(0.013)</td>
<td>(0.007)</td>
<td>(0.007)</td>
<td>(0.065)</td>
<td>(0.101)</td>
</tr>
<tr>
<td>Meeting #2</td>
<td>-0.188***</td>
<td>0.002</td>
<td>-0.019**</td>
<td>-0.017**</td>
<td>-0.030</td>
<td>-0.011</td>
</tr>
<tr>
<td>Meeting #3</td>
<td>-0.425***</td>
<td>-0.038***</td>
<td>-0.020**</td>
<td>-0.019**</td>
<td>-0.121**</td>
<td>-0.101**</td>
</tr>
<tr>
<td>Observations</td>
<td>1782</td>
<td>1782</td>
<td>1782</td>
<td>1782</td>
<td>118</td>
<td>118</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.50</td>
<td>0.31</td>
<td>0.21</td>
<td>0.21</td>
<td>0.07</td>
<td>0.05</td>
</tr>
<tr>
<td>Mean dep. variable</td>
<td>1.18</td>
<td>0.07</td>
<td>0.02</td>
<td>0.02</td>
<td>0.08</td>
<td>0.08</td>
</tr>
<tr>
<td>P-value Invitations = Invitations + Comment Forms</td>
<td>0.61</td>
<td>0.87</td>
<td>0.01</td>
<td>0.03</td>
<td>0.09</td>
<td>0.14</td>
</tr>
</tbody>
</table>

Notes: Data is from meeting survey. Each observation represents one village. “Serious response” is defined as agreeing to replace a supplier or village office, agreeing that money should be returned, agreeing for an internal village investigation, asking for help from district project officials, or requesting an external audit. “Serious response except audit” includes all of the actions except the last two. Estimation is by OLS. Robust standard errors in parentheses, adjusted for clustering by village (columns 1-4) or subdistrict (column 5-6).

* significant at 10%; ** significant at 5%; *** significant at 1%
Table 7: Participation -- Main theft results

Panel A: Invitations

<table>
<thead>
<tr>
<th>Percent missing:</th>
<th>Control Mean</th>
<th>Treatment Mean: Invite Effect</th>
<th>No Fixed Effects</th>
<th>Engineer Fixed Effects</th>
<th>Stratum Fixed Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log reported value – Log actual value</td>
<td></td>
<td></td>
<td>P-Value Invite Effect</td>
<td>P-Value Invite Effect</td>
<td>P-Value Invite Effect</td>
</tr>
<tr>
<td>Major items in roads</td>
<td>0.254 (0.035)</td>
<td>0.260 (0.035)</td>
<td>0.008 (0.035)</td>
<td>0.820 (0.034)</td>
<td>0.002 (0.034)</td>
</tr>
<tr>
<td>Major items in roads and ancillary projects</td>
<td>0.282 (0.032)</td>
<td>0.251 (0.032)</td>
<td>-0.028 (0.036)</td>
<td>0.431 (0.036)</td>
<td>-0.029 (0.036)</td>
</tr>
<tr>
<td>Breakdown of roads:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Materials</td>
<td>0.186 (0.042)</td>
<td>0.249 (0.042)</td>
<td>0.064 (0.043)</td>
<td>0.137 (0.041)</td>
<td>0.055 (0.035)</td>
</tr>
<tr>
<td>Unskilled labor</td>
<td>0.407 (0.088)</td>
<td>0.175 (0.088)</td>
<td>-0.231 (0.106)</td>
<td>0.032 (0.106)</td>
<td>-0.262 (0.106)</td>
</tr>
</tbody>
</table>

Panel B: Invitations + Comments

<table>
<thead>
<tr>
<th>Percent missing:</th>
<th>Control Mean</th>
<th>Treatment Mean: Invite + Comment Effect</th>
<th>No Fixed Effects</th>
<th>Engineer Fixed Effects</th>
<th>Stratum Fixed Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log reported value – Log actual value</td>
<td></td>
<td></td>
<td>P-Value Invite + Comment Effect</td>
<td>P-Value Invite + Comment Effect</td>
<td>P-Value Invite + Comment Effect</td>
</tr>
<tr>
<td>Major items in roads</td>
<td>0.254 (0.035)</td>
<td>0.255 (0.029)</td>
<td>0.002 (0.032)</td>
<td>0.950 (0.031)</td>
<td>0.002 (0.031)</td>
</tr>
<tr>
<td>Major items in roads and ancillary projects</td>
<td>0.282 (0.032)</td>
<td>0.245 (0.029)</td>
<td>-0.035 (0.035)</td>
<td>0.326 (0.035)</td>
<td>-0.035 (0.035)</td>
</tr>
<tr>
<td>Breakdown of roads:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Materials</td>
<td>0.186 (0.042)</td>
<td>0.168 (0.037)</td>
<td>-0.018 (0.034)</td>
<td>0.598 (0.033)</td>
<td>-0.011 (0.033)</td>
</tr>
<tr>
<td>Unskilled labor</td>
<td>0.407 (0.088)</td>
<td>0.308 (0.078)</td>
<td>-0.098 (0.093)</td>
<td>0.296 (0.095)</td>
<td>-0.125 (0.095)</td>
</tr>
</tbody>
</table>

Notes: Robust standard errors in parentheses. P-values from a regression of the dependent variable on invitations and comment form dummy variables using robust standard errors. Regressions without stratum (i.e., subdistrict) fixed effects include a variable for audits, and allow for clustering of standard errors by subdistrict.
Table 8: Invitations: Actual vs. Reported

<table>
<thead>
<tr>
<th>Log value – Log planned value</th>
<th>Control Mean</th>
<th>Treatment Mean: Invites</th>
<th>No Fixed Effects</th>
<th>P-Value</th>
<th>Engineer Fixed Effects</th>
<th>P-Value</th>
<th>Stratum Fixed Effects</th>
<th>P-Value</th>
<th>Num Obs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Materials</td>
<td>-0.221</td>
<td>-0.203</td>
<td>0.014</td>
<td>0.734</td>
<td>0.013</td>
<td>0.746</td>
<td>-0.003</td>
<td>0.932</td>
<td>406</td>
</tr>
<tr>
<td></td>
<td>(0.039)</td>
<td>(0.039)</td>
<td>(0.042)</td>
<td>(0.039)</td>
<td>(0.040)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unskilled labor</td>
<td>-0.295</td>
<td>-0.221</td>
<td>0.069</td>
<td>0.452</td>
<td>0.103</td>
<td>0.251</td>
<td>0.034</td>
<td>0.657</td>
<td>406</td>
</tr>
<tr>
<td></td>
<td>(0.080)</td>
<td>(0.080)</td>
<td>(0.092)</td>
<td>(0.090)</td>
<td>(0.077)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reported:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Materials</td>
<td>-0.041</td>
<td>0.019</td>
<td>0.059</td>
<td>0.004</td>
<td>0.055</td>
<td>0.002</td>
<td>0.035</td>
<td>0.160</td>
<td>406</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.013)</td>
<td>(0.020)</td>
<td>(0.018)</td>
<td>(0.025)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unskilled labor</td>
<td>0.063</td>
<td>-0.057</td>
<td>-0.124</td>
<td>0.115</td>
<td>-0.119</td>
<td>0.123</td>
<td>-0.175</td>
<td>0.099</td>
<td>406</td>
</tr>
<tr>
<td></td>
<td>(0.068)</td>
<td>(0.068)</td>
<td>(0.078)</td>
<td>(0.077)</td>
<td>(0.066)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>0.014</td>
<td>0.083</td>
<td>0.071</td>
<td>0.345</td>
<td>0.051</td>
<td>0.470</td>
<td>0.011</td>
<td>0.860</td>
<td>406</td>
</tr>
<tr>
<td></td>
<td>(0.068)</td>
<td>(0.068)</td>
<td>(0.075)</td>
<td>(0.070)</td>
<td>(0.064)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: See notes to Table 7. Dependent variable is the log of the actual or reported variable (as indicated) less the log of the planned value, i.e., the amount according to the initial budget, drawn up before the randomization occurred. Sample is limited to observations where all 5 dependent variables are non-missing.

Table 9: Net benefits calculation

<table>
<thead>
<tr>
<th></th>
<th>Audits</th>
<th>Invitations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Equal Weighted Net Benefits</td>
<td>Distribution Weighted Net Benefits</td>
</tr>
<tr>
<td>Cost of treatment:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monetary cost</td>
<td>-335</td>
<td>-278</td>
</tr>
<tr>
<td>Associated dead-weight loss</td>
<td>-134</td>
<td>-111</td>
</tr>
<tr>
<td>Time cost</td>
<td>-31</td>
<td>-31</td>
</tr>
<tr>
<td>Subtotal</td>
<td>-500</td>
<td>-419</td>
</tr>
<tr>
<td>Change in rents received by corrupt officials:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>From theft of materials</td>
<td>-224</td>
<td>-137</td>
</tr>
<tr>
<td>From theft of wages</td>
<td>-90</td>
<td>-55</td>
</tr>
<tr>
<td>Subtotal</td>
<td>-314</td>
<td>-192</td>
</tr>
<tr>
<td>Change in benefits from project:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPV value of road</td>
<td>1346</td>
<td>1346</td>
</tr>
<tr>
<td>Wages received by workers</td>
<td>123</td>
<td>123</td>
</tr>
<tr>
<td>Subtotal</td>
<td>1469</td>
<td>1469</td>
</tr>
<tr>
<td>TOTAL NET BENEFITS</td>
<td>655</td>
<td>858</td>
</tr>
</tbody>
</table>

Note: All figures in USD. Costs are listed as negative numbers. Distributional weights calculated using CRRA utility with coefficient of relative-risk aversion of 2. Derivation of change in NPV value of road and additional assumptions are discussed in the text.
Table 10: Assumptions for loss ratios and worker capacity

<table>
<thead>
<tr>
<th></th>
<th>Results from Calibration</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Loss Ratios</strong></td>
<td></td>
</tr>
<tr>
<td>Sand</td>
<td>1.00</td>
</tr>
<tr>
<td>Rocks</td>
<td>1.20</td>
</tr>
<tr>
<td>Gravel</td>
<td>1.75</td>
</tr>
<tr>
<td><strong>Worker Capacity</strong></td>
<td></td>
</tr>
<tr>
<td>Clearing brush and cleaning road surface (m²)</td>
<td>20</td>
</tr>
<tr>
<td>Spreading sand (m³)</td>
<td>4.5</td>
</tr>
<tr>
<td>Splitting rocks (m³)</td>
<td>3.0</td>
</tr>
<tr>
<td>Installing rock layer (m²)</td>
<td>6.5</td>
</tr>
<tr>
<td>Spreading gravel (m³)</td>
<td>2.25</td>
</tr>
<tr>
<td>Digging side channels and creating shoulder (m³)</td>
<td>1.0</td>
</tr>
<tr>
<td>Building retaining wall (m³) (unskilled labor portion)</td>
<td>0.28*</td>
</tr>
<tr>
<td>Building retaining wall (m³) (skilled labor portion)</td>
<td>0.83*</td>
</tr>
<tr>
<td>Standard cut and fill(m³)</td>
<td>2*</td>
</tr>
</tbody>
</table>

Notes: All loss ratios are expressed as the ratio of the original amount of material to the amount measured, and are for measurements loose, not compacted. Worker capacity is the quantity of the given activity that can be done by one person per 6 hours of work.
* Original KDP assumption that was not able to be reconfirmed from calibration exercises.

Figure 1: Relationship between theft of wages and percent of workers from outside village

Notes: Predicted values from a Fan locally-weighted regression with quartic kernel using bandwidth of 50. Confidence intervals obtained from a bootstrap with 200 replications.