Monitoring Corruption: Evidence from a Field Experiment in Indonesia

Author(s): Benjamin A. Olken

Source: *Journal of Political Economy*, Vol. 115, No. 2 (April 2007), pp. 200-249

Published by: The University of Chicago Press

Stable URL: https://www.jstor.org/stable/10.1086/517935

Monitoring Corruption: Evidence from a Field Experiment in Indonesia

Benjamin A. Olken

Harvard University and National Bureau of Economic Research

This paper presents a randomized field experiment on reducing corruption in over 600 Indonesian village road projects. I find that increasing government audits from 4 percent of projects to 100 percent reduced missing expenditures, as measured by discrepancies between official project costs and an independent engineers' estimate of costs, by eight percentage points. By contrast, increasing grassroots participation in monitoring had little average impact, reducing missing expenditures only in situations with limited free-rider problems and limited elite capture. Overall, the results suggest that traditional top-down monitoring can play an important role in reducing corruption, even in a highly corrupt environment.

I. Introduction

Corruption is a significant problem 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

I wish to thank Alberto Alesina, Abhijit Banerjee, Robert Barro, Dan Biller, Stephen Burgess, Francesco Caselli, Joe Doyle, Esther Duflo, Pieter Evers, Amy Finkelstein, Brian Jacob, Seema Jayachandran, Ben Jones, Larry Katz, Philip Keefer, Michael Kremer, Jeff Liebman, Erzo Luttmer, Ted Miguel, Chris Pycroft, Lant Pritchett, Mike Richards, Mark Rosenzweig, numerous seminar participants, Steven Levitt (the editor), and two anonymous referees 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 Department for International Development–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.

[*Journal of Political Economy*, 2007, vol. 115, no. 2]

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 best to 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 monitoring and enforcing 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 top-down 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 grassroots participation by community members in local-level monitoring. 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 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, 1). The idea behind the grassroots 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. Grassroots monitoring may also be prone to capture by local elites (Bardhan 2002; Bardhan and Mookherjee 2006). Given these countervailing forces, whether grassroots monitoring can actually succeed in reducing corruption is also an empirical question.[3]

To examine these alternative approaches to fighting corruption, I

[1] See, e.g., 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), e.g., discusses this possibility and shows that this type of multitiered corruption can lead to multiple equilibria in corruption.

[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.

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 randomly 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.

To investigate the impact of increasing community participation in the monitoring process, I designed two different experiments that sought to increase grassroots monitoring of the project. Specifically, the experiments sought to enhance participation at "accountability meetings," the village-level meetings in which 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 in sealed drop boxes before the accountability meetings, and the results were summarized at the meetings. Both of these experimental interventions were successful in raising grassroots participation levels: the invitations increased the number of people participating in the accountability meetings by about 40 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.[4] This paper, however, builds on a small but growing literature that examines corruption by comparing two measures of the same quantity, one "before" and one "after" corruption

[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).

has taken place.[5] To do this in the context of 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. The difference between what the village claimed the road cost to build and what the engineers estimated it actually cost to build is the key measure of missing expenditures I examine in this paper. 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. Missing expenditures averaged about 24 percent across the villages in the study.[6]

I find that there were substantial reductions in missing expenditures associated with the audit experiment. In particular, I estimate that the audit treatment—that is, increasing the probability of an audit from a baseline of 4 percent to 100 percent—was associated with reductions in missing expenditures of about eight percentage points. These reductions came from reductions in both unaccounted-for materials procured for the project and unaccounted-for labor expenditures. Interestingly, I find that the number of project jobs given to family members of project officials actually increased in response to the audits, which provides suggestive evidence that alternative forms of corruption may be substitutes. I show that while the auditors' findings are positively correlated with the findings from my independent engineering survey, in the vast majority of cases the auditors' findings were procedural in nature, and not the sort of "caught-red-handed" evidence that could be used to prove

[5] For example, Reinikka and Svensson (2004) examine corruption in educational expenditures, Fisman and Wei (2004) and Yang (2004) examine corruption in international trade, Di Tella and Schargrodsky (2003) examine corruption in hospital procurement, and Olken (2006*a*) examines theft from a government redistribution program. This paper differs 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. In that sense, it is related to the paper by Hsieh and Moretti (2006), who compare prices received by Iraq under the U.N. Oil-for-Food program to the world oil price.

[6] These "missing expenditures," i.e., the difference between reported expenditures and my estimate of actual expenditures, may also include sources of losses other than pure theft. I discuss below how I constructed several test roads to estimate the typical amount of materials lost during construction; I use them to calibrate the missing expenditures measure to be zero in a road for which I know a priori that there was no corruption. I also discuss below how I use independent measures of the quality of road construction that are likely to be unrelated to corruption to differentiate between overall changes in the competence of road builders and corruption per se.

criminal malfeasance. This may help explain why almost 20 percent of expenditures were still unaccounted for even in villages facing a 100 percent probability of an external government audit.

By contrast, I find that the participation experiments—the invitations and the anonymous comment forms—were associated with much smaller, and statistically insignificant, average reductions in overall missing expenditures. The idea behind community monitoring is that while the village implementation team has incentives to steal from the project, the village at large would benefit from the higher road quality associated with less corruption. 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 missing expenditures.

The small overall effects of the participation experiments on overall missing expenditures, however, mask substantial differences across types of expenditures and different ways of implementing increased grassroots participation. In particular, the invitations treatment substantially reduced missing labor expenditures but had no effect whatsoever on missing materials expenditures.[7] I present suggestive evidence that the reason for the differential effect on labor and materials is that community members had a strong incentive to monitor wage payments, whereas free-riding was much more of a problem for materials expenditures. I also show that the anonymous comment form treatment did reduce missing expenditures in some cases, but only when the comment forms were distributed entirely via village schools, completely bypassing the village government and preventing village elites from disproportionately channeling the forms to their supporters. These results suggest that while grassroots monitoring has the potential to reduce corruption, care must be taken to minimize free-rider problems and prevent elite capture.

The remainder of the paper is organized as follows. Section II discusses the setting in which the study takes place. Section III describes the experimental interventions. Section IV describes the data used in the study. Section V presents the results of the experiments. Section VI performs a cost-benefit analysis. Section VII presents conclusions.

[7] Since materials account for about three-quarters of total expenditures, the average impact on missing expenditures was small and statistically insignificant.

II. Setting

The Kecamatan (subdistrict) Development Project, or KDP, is a national Indonesian 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, and were collected between September 2003 and August 2004.

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 percent) propose an infrastructure project plus a small amount for savings and loans for women. An intervillage forum ranks all 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.

If the project is funded, a meeting is held to plan construction, after which an elected implementation team procures materials, hires labor, and builds the project. 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.[8] No contractors are used in construction.

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 annual 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 to 3 kilometers and may either run within the village or run from the village to the fields. Dirt roads in Java are typically impassable during the rainy

[8] 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).

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 nonasphalt road projects.

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 percent, 40 percent, and 20 percent 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 attended by only 30–50 people, most of whom are members of the village elite, out of an average village adult population of about 2,500.

In addition, subdistrict-, district-, and provincial-level project managers, engineers, and facilitators conduct overall supervision of all projects, and there is a provincial complaints-handling unit that 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 percent baseline chance of being audited by BPKP. If the village is selected for an audit, auditors come to the village, cross-check all the financial records looking for irregularities, and 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 of village-level officials are rare in practice (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.[9] 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 kickback to village and project officials. Second, members of the implementation team may manipulate wage payments. As discussed above, villagers in Indonesia typically contribute unpaid or reduced-wage labor to public

[9] Of course, there may also be collusion or kickbacks at the national or district level of the program. This paper, however, focuses on corruption in which the bulk of the program money is actually spent—at the village level.

TABLE 1
NUMBER OF VILLAGES IN EACH TREATMENT CATEGORY

	Control	Invitations	Invitations Plus Comment Forms	Total
Control	114	105	106	325
Audit	93	94	96	283
Total	207	199	202	608

NOTE.—Tabulations are taken from results of the randomization. Each subdistrict faced a 48 percent chance of being randomized into the audit treatment. Each village faced a 33 percent chance of being randomized into the invitations treatment and a 33 percent chance of being randomized into the invitations plus comment forms treatment. The randomization into audits was independent of the randomization into invitations or invitations plus comment forms.

works projects; in such cases, corrupt officials can bill the project for the voluntary labor anyway and pocket the difference. In other cases, those running the project can simply inflate the number of workers paid by the project. All these types of corruption will be investigated in the empirical work below.

III. 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 ("invitations plus comments"). Section III.*A* discusses the overall experimental design. Section III.*B* then discusses the audit interventions, and Section III.*C* describes the invitations and comment interventions. Section III.*D* discusses the timing of the interventions and data collection.

A. *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.[10] Thus the choice of what type of

[10] In all villages (including control villages), at the village meeting immediately after the final allocations were announced but before construction began, the study enumerator made a short (less than five-minute) presentation, introducing himself 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

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/comments treatments differed in several ways. First, there was a concern that the audit treatment might be likely to spill over from one village to another, since officials in other villages might worry that when the auditors came to the subdistrict, their villages might be audited as well.[11] On the other hand, the participation treatments were much less likely to have similar spillover effects, since 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), whereas 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 that 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 including stratum fixed effects, while it may reduce standard errors, is not necessary for the analysis to be consistent.

Although locations of treatments were randomized by computer ac-

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.

[11] This was most likely to be a problem within subdistricts, since 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. In results not reported, to test for the presence of spillovers across subdistricts, for all villages not in the audit treatment I calculate the distance to the nearest audit village. I find no impact of this distance variable on missing expenditures, which suggests that indeed these cross-subdistrict spillovers are minimal.

TABLE 2
RELATIONSHIP BETWEEN TREATMENTS AND VILLAGE CHARACTERISTICS

	Audits (1)	Invitations (2)	Comments (Conditional on Invitations) (3)
Village population (000s)	−.007	.004	.001
	(.012)	(.007)	(.009)
Mosques per 1,000	−.018	.000	.012
	(.038)	(.024)	(.028)
Total budget (Rp. millions)	−.001	−.000	−.000
	(.001)	(.000)	(.000)
Number subprojects	−.017	.002	−.017
	(.025)	(.013)	(.016)
Percent households poor	.246*	.069	.033
	(.126)	(.080)	(.111)
Distance to subdistrict	−.001	−.002	.001
	(.005)	(.004)	(.005)
Village head education	.012	−.002	.016
	(.009)	(.007)	(.010)
Village head age	.004	.003	−.000
	(.003)	(.003)	(.003)
Village head salary (hectares)	.011*	.004	.006
	(.007)	(.003)	(.004)
Mountainous dummy	.134*	.010	.077
	(.074)	(.037)	(.049)
Observations	577	577	381
p-value of all listed variables	.18	.92	.52

NOTE.—Results reported are marginal effects from Probit regressions. Robust standard errors are in parentheses, adjusted for clustering at the subdistrict level.
* Significant at 10 percent.
** Significant at 5 percent.
*** Significant at 1 percent.

cording to the procedures described above, it is useful to examine whether, ex post, they are correlated with village characteristics of interest. Table 2 examines this by reporting the results of Probit regressions of the probability of being randomized into each treatment group on 10 village characteristics.[12] As expected given the randomization, these variables are not jointly significant predictors of the treatments (joint p-value $= 0.18$ for audits, 0.92 for invitations, and 0.52 for comments conditional on invitations), though several variables (village poverty rate, village head salary, and a dummy for being in a mountainous area) are individually significant at the 10 percent level in the audit equation. However, the main results in the paper do not change substantially if I include all these controls as explanatory variables (results available from the author on request).

[12] The variables examined are the variables used in analysis of pilot data and were specified before any of the data used here were collected.

B. *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 one. They were told that the audit could take place either during or after construction was finished and would include both inspections of the project's financial records and a field inspection of the construction activities. Approximately two months later, the village implementation team received a one-page letter from BPKP that confirmed that the village had been chosen to be audited and 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.

Between one and four months after construction had started, phase I of the audits commenced.[13] The main purpose of this first round of audits was to credibly demonstrate that the 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 public village meeting, where 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 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 follow-up.

[13] 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 one and four months after construction had begun.

C. *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.[14] 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 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 heads) were randomized by village. 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.

In the invitations plus comment forms treatment, invitations were distributed exactly as in the invitations treatment, but attached to the invitation was a comment form asking villagers' opinions of the project. The idea behind the comment form was that villagers might be afraid of retaliation from village elites, and thus providing an anonymous comment form would increase detection of corruption. 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 subvillage. The instructions stated clearly that the recipients should not write their names on the form, in order to preserve their anonymity. According to the household survey conducted as part of the project, 89 percent of adults in these villages can read and write, which suggests that literacy is sufficient for most villagers to fill out the form. The form had three closed-response questions (i.e., requesting answers of the form good,

[14] 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.

satisfactory, or poor) about various aspects of the project and two free-response questions, one asking about the job performance of the implementation team and one asking about any other project-related issues. The comment forms were collected from the drop boxes two days before each meeting and summarized by a project enumerator. The enumerator then read the summary, including a representative sample of the open-response questions, at the village meeting.

D. Timing

The experiment began in September 2003. After the intervillage forum described in Section II made the final allocations of funds, the enumerator went to the village planning meeting that immediately followed and, at that planning meeting, announced any interventions (audits or participation) that would take place in that village. Construction began shortly thereafter, between October and November 2003. Those villages receiving audits received the first letter from BPKP in November 2003, and the first round of audits took place in one randomly selected village in each subdistrict in January 2004, while construction was in progress. The second letter from BPKP was sent out to villages shortly thereafter. The accountability meetings at which the participation interventions were conducted took place after 40 percent, 80 percent, and 100 percent of the funds were spent, between October 2003 and May 2004. The engineering survey to measure missing expenditures took place after construction was finished, between May and August 2004. The final round of audits was conducted in all villages in the audit treatment in September 2004 after all of the data collection for the engineering survey had been completed.

IV. 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 engineering field survey, used to measure corruption in the project. This measurement was conducted in all villages (both treatment and control) and is completely separate from the audits conducted by BPKP as part of the audit treatment. This section describes the final field survey used to measure unaccounted-for expenditures in the road projects; the remaining data, as well as additional details on the field survey, are discussed in more detail in Appendix A.

A. *Reported Expenditures*

The key dependent variable I examine is the difference between what villages claim they spent on the project and an independent estimate of what villages actually spent. Obtaining data on 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.

B. *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 centimeters) rocks, and a top layer of gravel to provide a smooth running surface.[15] To estimate the quantity of each of these materials, the engineers dug 10 40 centimeter × 40 centimeter 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

[15] Three other similar road designs are also included in the study. *Telasah* roads are similar to telford but the rocks are installed flat side up to create a smooth running surface, and therefore the gravel layer is largely omitted. *Sirtu* roads consist of gravel only, with the sand and rock layer omitted. *Katel* roads are similar in design to telford but contain 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. The type of road is chosen before the randomization is announced.

of the total length and average width of the road, I can estimate the total quantity of materials used in the road.

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, since 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 as a result of normal construction processes and measurement error but with no corruption as the *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 one is combining different types of materials into the aggregate percent missing measure, 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 the percent missing variable, not just the differences across villages.

To obtain such an estimate, I constructed four short (60-meter) "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). After construction was completed, the techniques described above were used several times, by different engineers, to estimate the quantity of materials used in the road. To allow time for materials to settle and to account for the effects of weather, these follow-up measurements were conducted anywhere from one week to one year after the test road was completed. 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 constitutes 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.

C. 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, and what the daily wage and number of hours worked were. They were also asked 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 from both 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.

D. 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 as recent buyers of material (primarily workers at construction sites).[16] For each type of material used by the project, between three and five independent prices were obtained; I use the median price from the survey for the analysis.

E. Measures of Missing Expenditures

From village financial reports, I calculate the total expenditures the village claimed it incurred in building the project, which I hereafter refer to as the *reported* amount. From the field survey, I estimate the total expenditures the village actually incurred in building the project, which I hereafter refer to as the *actual* amount. I define the *percent missing* to be the difference between the log of the reported amount

[16] 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 percent of those interviewed for the price survey had actually been suppliers to the KDP program; dropping them from the 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.

TABLE 3
Summary Statistics

	Summary Statistics
Total project size (US$)	8,875 (4,401)
Share of total reported expenses:	
Road project	.766 (.230)
Ancillary projects (culverts, retaining walls, etc.)	.154 (.181)
Other projects (schools, bridges, irrigation, etc.)	.079 (.166)
Share of reported road expenses:	
Sand	.099 (.080)
Rocks	.484 (.143)
Gravel	.116 (.181)
Unskilled labor	.196 (.125)
Other	.105 (.164)
Percent missing:	
Major items in road project	.237 (.343)
Major items in roads and ancillary projects	.247 (.350)
Materials in road project	.203 (.395)
Unskilled labor in road project	.273 (.851)
Observations	538

Note.—Statistics shown are means, with standard deviations in parentheses. Data on expenditures are taken from the 538 villages for which percent missing in road and ancillary projects could be calculated. Exchange rate is Rp. 9,000 = US$1.00.

and the log of the actual amount. This variable—the percent missing—is the main measure used in the subsequent analysis.

I use several different versions of the percent missing measure in the empirical analysis. First, I report the percent missing for the four major items—sand, rocks, gravel, and unskilled labor—used in the road project. As shown in table 3, expenditures on these four items account for 90 percent 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.

As shown in table 3, the road project accounts for 77 percent of total funds spent; a further 15 percent of funds is 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 missing expenditures, major items in the main road plus ancillary projects, adds in these expenditures as well.[17] Finally, I report the percent missing separately for materials and unskilled labor in the road project.

[17] 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 corruption measure could not be constructed for just the main road, but it could be computed for the main road plus ancillary projects.

Table 3 reports the average values for the four different percent missing variables used in the study. The figures show that, on average across the sample, about 24 percent of expenditures could not be accounted for. This figure is similar in magnitude to that found in other studies, such as the second-round Public Expenditure Tracking Survey in Uganda reported in Reinikka and Svensson (2005) and an estimate of unaccounted-for rice in a subsidized food program in Indonesia reported by Olken (2006*a*). However, the absolute levels of the percent missing variable depend on assumptions for loss ratios and should therefore be interpreted with caution. To the extent that the calibration exercises for the loss ratios did not fully capture actual loss levels, the true levels of corruption could be somewhat lower than the averages reported here.

V. Experimental Results

A. Estimating Equation

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

$$\text{PercentMissing}_{ijk} = \alpha_1 + \alpha_2\text{Audit}_{jk} + \alpha_3\text{Invitations}_{ijk} + \alpha_4\text{InvitationsandComments}_{ijk} + \epsilon_{ijk}, \qquad (1)$$

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 (1) that includes engineering team fixed effects. Finally, when investigating the audits, I estimate a version of equation (1) that includes fixed effects for each audit stratum k, and when investigating the invitations and comment forms, I estimate a version of equation (1) that includes fixed effects for each subdistrict j (i.e., the stratifying variable for the participation experiments).[18]

[18] Note that approximately 12 percent 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 occurred for one of four reasons: (1) surveyor error in locating the road, (2) a project consisting 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. A regression of the village being dropped for any of these reasons on the three treatment dummies reveals that being absent from the sample is orthogonal to the treatments.

B. *The Audit Experiment*

1. Overall Effects

Table 4 presents the main results from the audit experiment. Each row presents the percent missing in different aspects of the project. Column 1 presents the mean percent missing in the control villages—that is, those villages that did not receive the audits—and column 2 presents the mean level in the villages that received the audits. The effect of the audits—that is, the coefficient α_2 in equation (1)—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 the percentage of expenditures that could not be accounted for. Column 3 shows that the audits reduced the percent missing in the road project by 8.5 percentage points and the percent missing in the road and ancillary projects by 9.1 percentage points. These effects are statistically significant, with p-values of 0.058 and 0.034, respectively. In the other columns, these effects generally have a magnitude and statistical significance similar to those of engineer or stratum fixed effects. The only exception is the case in which stratum fixed effects are included in the road expenditures variable (col. 7, row 1), in which case the estimated effect of the audits is only a 4.8-percentage-point reduction.[19] Figure 1 shows the results graphically, presenting for each group the empirical cumulative distribution functions (CDFs) of the percent missing variable and the estimated probability density functions (PDFs) of the percent missing variable, where the PDFs are estimated using kernel density methods. The results in figure 1 show that the percent missing in the audit group is first-order stochastically dominated by the percent missing in the control group, showing that the reduction in missing expenditures occurred at all percentiles of the distribution.

Looking across all the specifications shown, I conclude that the audits

[19] The reason that the results are different when stratum fixed effects are included is that doing so effectively removes 13 strata from the sample. 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. This, plus dropping observations for the reasons described in nn. 17 and 18, led to 13 out of 50 strata with either all audit or all nonaudit villages that are effectively dropped by stratum fixed effects. Estimating the overall results dropping these 13 strata, but without any fixed effects, yields results similar to stratum fixed effects results. This problem is less severe for the major items in roads and ancillary projects, in part because fewer observations are missing and thus fewer strata are effectively dropped.

TABLE 4

AUDITS: MAIN THEFT RESULTS

			No Fixed Effects		Engineer Fixed Effects		Stratum Fixed Effects	
Percent Missing[a]	Control Mean (1)	Treatment Mean: Audits (2)	Audit Effect (3)	p-Value (4)	Audit Effect (5)	p-Value (6)	Audit Effect (7)	p-Value (8)
Major items in roads (N = 477)	.277	.192	−.085*	.058	−.076**	.039	−.048	.123
	(.033)	(.029)	(.044)		(.036)		(.031)	
Major items in roads and ancillary projects (N = 538)	.291	.199	−.091**	.034	−.086**	.022	−.090***	.008
	(.030)	(.030)	(.043)		(.037)		(.034)	
Breakdown of roads:								
Materials	.240	.162	−.078	.143	−.063	.136	−.034	.372
	(.038)	(.036)	(.053)		(.042)		(.037)	
Unskilled labor	.312	.231	−.077	.477	−.090	.304	−.041	.567
	(.080)	(.072)	(.108)		(.087)		(.072)	

NOTE.—Audit effect, standard errors, and p-values are computed by estimating eq. (1), a regression of the dependent variable on a dummy for audit treatment, invitations treatment, and invitations plus comment forms treatments. Robust standard errors are in parentheses, allowing for clustering by subdistrict (to account for clustering of treatment by subdistrict). Each audit effect, standard error, and accompanying p-value is taken from a separate regression. Each row shows a different dependent variable, shown at left. 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 missing. Villages are included in each row only if there was positive reported expenditures for the dependent variable listed in that row.

[a] Percent missing equals log reported value − log actual value.

* Significant at 10 percent.

** Significant at 5 percent.

*** Significant at 1 percent.

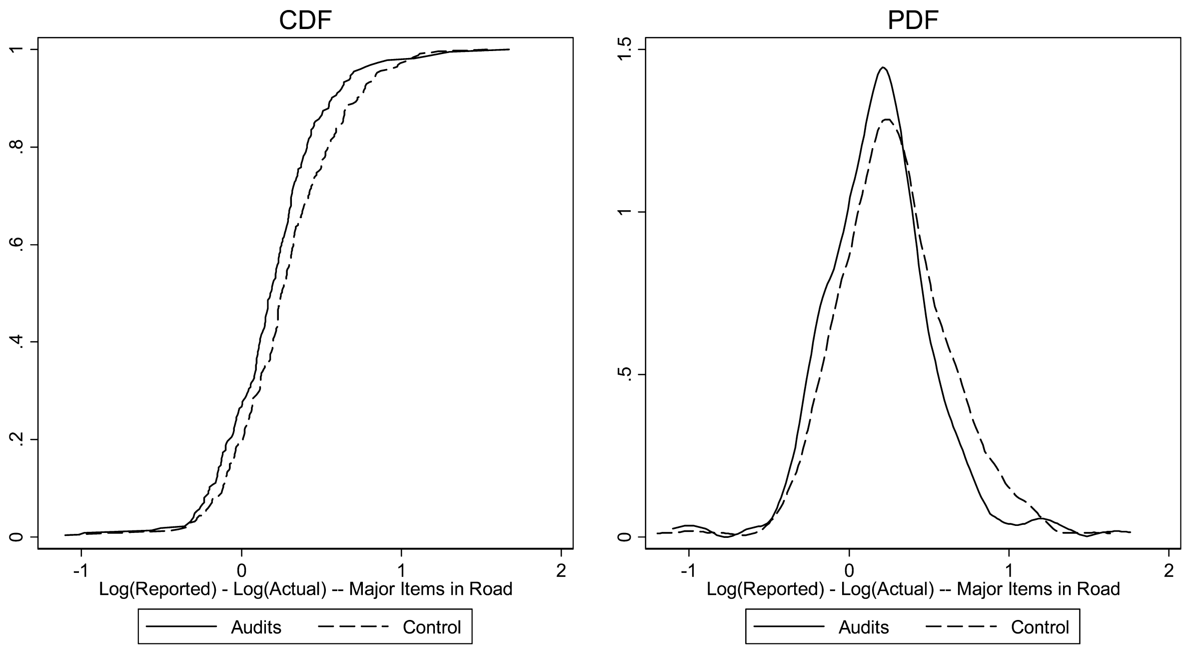

Fig. 1.—Empirical distribution of missing expenditures. The left-hand figure shows the empirical CDF of missing expenditures for the major items in a road project, separately for villages in the audit treatment group (solid line) and the control group (dashed line). The right-hand figure shows estimated PDFs of missing expenditures for both groups; PDFs are estimated using kernel density regressions using an Epanechnikov kernel.

reduced missing expenditures by an average of about eight percentage points. Compared with a level of 27.7 percentage points in control villages, the point estimates imply a reduction in missing expenditures of about 30 percent of the level in control villages, although, as discussed above, the absolute levels of the percent missing variable depend on assumptions for loss ratios and should be interpreted with caution.

Breaking down the change in percent missing into materials and labor, table 4 shows substantial reductions in both materials (sand, rocks, and gravel) and unskilled labor associated with the audits, though these separate effects are not statistically significant.[20] Interestingly, in results not reported in the table, I find no significant differences in the effect of the audits between those villages audited both during and after construction and those villages audited only after construction was finished (and, therefore, after the engineering survey was completed). This suggests that the reduction in missing expenditures was caused by the threat of an audit rather than corrective actions imposed by the auditors once they arrived.

Mechanically, unaccounted-for funds must be accounted for by either differences in the price charged per unit or differences in the quantities used. When the goods being procured are commodities (as they are in this case), it is much easier for monitors to verify the unit price than to verify the quantity of materials used, so one might expect corruption to occur by inflating quantities rather than prices. To investigate this, in table 5, I decompose the results into differences in prices and differences in quantities. Column 1 shows that even in control villages, there is almost no difference between reported and actual prices, and that all the unaccounted-for expenditures were due to differences between reported and actual quantities.[21] Consistent with there being no markups of prices to begin with, columns 3–8 show that all the reductions in missing expenditures caused by the audits were on the quantity dimension.

An important question is whether the observed effects of the audits actually represent a reduction in corruption per se or whether incompetent builders are simply being replaced by more skilled builders in response to the audits. To investigate this, I examined 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 at construction should affect both expensive and inexpensive components, whereas a reduction in corruption would disproportionately affect the expensive aspects of construction

[20] It is worth noting that, because of the log transformation, mechanically the change in log(materials + wages) and log(materials) + log(wages) will not be identical.

[21] This difference is statistically significant: a t-test in control villages rejects the equality of differences in prices and differences in quantities with a p-value of <0.01.

TABLE 5

AUDITS: PRICES VS. QUANTITIES

PERCENT MISSING[a]	CONTROL MEAN (1)	TREATMENT MEAN: AUDITS (2)	NO FIXED EFFECTS: Audit Effect (3)	NO FIXED EFFECTS: p-Value (4)	ENGINEER FIXED EFFECTS: Audit Effect (5)	ENGINEER FIXED EFFECTS: p-Value (6)	STRATUM FIXED EFFECTS: Audit Effect (7)	STRATUM FIXED EFFECTS: p-Value (8)
Prices (N = 494)	−.018	−.016	.002	.949	.011	.662	.018	.434
	(.017)	(.021)	(.027)		(.025)		(.023)	
Quantities (N = 477)	.276	.207	−.069*	.092	−.069*	.051	−.048	.120
	(.030)	(.028)	(.041)		(.035)		(.031)	

NOTE.—See the note to table 4. Reported and actual prices are defined as a weighted average across different commodities (rock, sand, gravel, and unskilled labor), where each commodity is weighted by the reported quantity. Reported and actual quantities are also defined as a weighted average, where each commodity is weighted by the reported prices.

[a] Percent missing equals log reported value − log actual value.

* Significant at 10 percent.

** Significant at 5 percent.

*** Significant at 1 percent.

quality (i.e., the volume of materials). In fact, I find that controlling for these inexpensive quality measures, either individually or aggregated into an index, does not change the corruption results presented above.[22] This suggests that the results are actually being driven by a reduction in corruption per se rather than an overall change in the competence of those building the project.

Another important question is whether the changes in corruption represent a change in actual expenditures or simply reflect more careful accounting by villages in response to the upcoming audits. To examine this, 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 are 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 from the final expenditure report and the log of the corresponding amounts from the initial planned budget). In results not reported in the table (but available from the author on request), I find that the changes associated with the audits were driven by increases in actual expenditures—for both materials and unskilled labor—rather than changes in reported expenditures, though the results are not statistically significant.

2. Examining Auditors' Findings

A natural question is why the impact of the audits was not larger; that is, the point estimates suggest that even with an audit probability of one, about 20 percent of funds were still not accounted for. One potential explanation is that even if the probability of being audited is one, the probability that corruption is detected and a punishment is imposed may still be much less than one. This subsection investigates several reasons why audits might not necessarily result in punishment of corrupt officials and concludes that while auditors are able to detect corruption, the evidence is often too circumstantial to form the basis of a prosecution.

First, to investigate the quality of the audits, I compare the results from the auditors' final reports to the results from the independent

[22] Alternatively, putting an index of these variables on the left-hand side of eq. (1) shows that there is no change in these inexpensive quality measures in response to the audits.

TABLE 6
Relationship between Auditor Findings and Survey Team Findings

	Engineering Team Physical Score (1)	Engineering Team Administrative Score (2)	Percent Missing in Road Project (3)
Auditor physical score	.109**	−.067	.024
	(.043)	(.071)	(.033)
Auditor administrative score	.007	.272**	−.055**
	(.049)	(.133)	(.027)
Subdistrict fixed effects	Yes	Yes	Yes
Observations	248	249	212
R^2	.83	.78	.46

Note.—Robust standard errors are in parentheses, adjusted for clustering at subdistrict level. Auditor scores refer to the results from the final BPKP audits; engineering team scores refer to the results from the engineering team that was sent to estimate missing expenditures. The results from the engineering team were not shared with the BPKP audit team. All specifications include subdistrict fixed effects, which therefore hold constant both the BPKP audit teams and the engineering teams. For both physical and administrative scores, scores are normalized to have mean zero and standard deviation one.
* Significant at 10 percent.
** Significant at 5 percent.
*** Significant at 1 percent.

engineering survey.[23] Specifically, the auditors filled out a long checklist on project infrastructure quality and administrative issues, rating each checklist item on a three-point scale (satisfactory, deficient, very deficient). I normalize the average finding on the infrastructure and administrative checklists, denoted Auditor Physical Score and Auditor Admin Score, respectively, to each have mean zero and standard deviation one, with higher numbers indicating a better score.

At the time of the independent field survey, the engineers filled out an identical checklist, in addition to collecting the data used to construct the missing expenditures variable. In table 6, I investigate the relationship between the scores from the auditors' checklists and the analogous measures from the engineering team. I estimate the following regression:

$$\text{EngineeringScore}_{ij} = \alpha_j + \beta_1 \text{AuditorPhysicalScore}_{ij} + \beta_2 \text{AuditorAdminScore}_{ij} + \epsilon_{ij}, \qquad (2)$$

where i represents a village and j represents a subdistrict. The inclusion of subdistrict fixed effects holds constant both the BPKP auditing team and the engineering team and thus captures average differences in how different teams filled out the checklist. The results in table 6 show that the physical score given by BPKP is positively correlated with the physical score given by the engineering team from my survey (col. 1); similarly,

[23] The information collected by the engineering team was not shared with the audit team. In fact, in the case of the missing expenditures measure, the survey team gathering data on missing expenditures collected raw data, such as the depth of surface layers; all processing to calculate missing expenditures was done subsequently by computer.

TABLE 7
AUDIT FINDINGS

	Percentage of Villages with Finding
Any finding by BPKP auditors	90%
Any finding involving physical construction	58%
Any finding involving administration	80%
Daily expenditure ledger not in accordance with procedures	50%
Procurement/tendering procedures not followed properly	38%
Insufficient documentation of receipt of materials	28%
Insufficient receipts for expenditures	17%
Receipts improperly archived	17%
Insufficient documentation of labor payments	4%

NOTE.—Tabulations from BPKP final report submitted to the Government of Indonesia's KDP management team and to the World Bank on December 22, 2004. This report included all findings from the 283 villages that were audited as part of phase II of the audits. The percentage reported is the percentage of the 283 audited village for which BPKP reported finding the listed problem.

the BPKP administrative score is positively correlated with the engineering team administrative score (col. 2). Even more important, column 3 shows that the BPKP administrative score is strongly negatively correlated with the missing expenditures measure. Specifically, a one-standard-deviation better score on the BPKP checklist is associated with 5.5 percentage points less missing expenditures. All told, these results suggest that the auditors were not completely corrupt (i.e., their results were correlated with the results from the independent engineering team) and that the administrative aspects investigated by the auditors were in fact correlated with missing expenditures.

A second potential reason why audits might not have led to punishments is that the problems they detect may not constitute sufficient evidence to impose a criminal punishment. To investigate this, table 7 tabulates the "findings" reported in the final audit reports from the second phase of audits. While auditors reported at least one finding in 90 percent of the villages they visited, most of these findings were that procedures had not been properly followed (e.g., the tendering process for procurement was not properly followed in 38 percent of villages, receipts were incomplete in 17 percent of villages, etc.) rather than concrete evidence of malfeasance.[24] Reports of such findings by BPKP

[24] For example, the finding that the tendering process for procurement was not followed might mean that "tenders were not submitted in writing, but instead were only submitted orally" (28) or that "the auditors could not locate price survey or tender documents" (26). The finding that receipts were insufficient might mean that "purchase of 300 sacks of Portland cement could not be verified because no receipt was present" (44–45) or that "reimbursement of operational expenses of Rp. 1,840,000 (US$200) to head of implementation team was not supported by receipts" (47). While a lack of receipts or lack of documentation from a tender process may be suspicious, it does not in itself constitute evidence of malfeasance.

might have created some discomfort for village officials, since they would need to provide explanations for the irregularities, but they might not have led to criminal sanctions even in the presence of a fully functioning legal system.

A final reason why the audits may not have had a larger impact is that the punishment conditional on exposing corruption may have been relatively weak. Given the low probability of formal criminal prosecution, an important part of the threat posed by an audit is that the village head might lose his reelection bid if corruption was exposed.[25] In results reported in the working paper version of this paper (Olken 2005), I find that the effect of the audits is most pronounced for village heads who are up for reelection in the next two years, who would face the largest effective punishment for corruption being detected; the point estimates suggest that missing expenditures would be reduced to 0.06 in audit villages in which elections were coming up in the next two years. This result suggests that auditing and punishments—and, specifically, punishments through democratic accountability—may be complements and, more generally, that audits may be more effective with higher punishments.

3. Employment of Family Members

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 family members is an alternative, but less desirable, method of extracting rents from the project, we might expect this type of nonaudited behavior to respond to the increased audits.[26] Of course, there are other reasons to hire family members besides rent extraction; for example, work on the project by family members might increase if project officials facing audits wanted to improve the project and if family members are less prone to moral hazard or have higher skills 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.

[25] As discussed above, criminal corruption proceedings are becoming more common in Indonesia, but prosecutions against village-level officials remain rare (Woodhouse 2004). Electoral and social sanctions against village-level officials are therefore likely to be a substantial portion of the threat created by audits. In fact, the village head's vote share in the previous election was negatively related to missing expenditures in the road project. This suggests that less corrupt officials may receive more votes, though other interpretations of this result are also possible.

[26] The increase in nepotism can be thought of as an increase in an inferior type of corruption (nepotism) in response to the negative income shock from the loss of rents due to the audits.

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. Overall, approximately 30 percent of respondents are related to some member of the village government, and 6 percent of respondents are related to the head of the project. I examine whether people who said that they were either immediate or extended family members of village government or project officials 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:[27]

$$\text{Worked}_{hijk} = \gamma_k + \gamma_2\text{Audit}_{jk} + \gamma_3\text{Family}_{hijk} + \gamma_4\text{Audit} \times \text{Family}_{ijk} + \gamma_5 \boldsymbol{X}_{hijk} + \epsilon_{hijk}, \tag{3}$$

where Family is a dummy equal to one if the individual was a family member of a village government official or the head of the project; Worked is a dummy equal to one if a household member worked for pay on the project; $\boldsymbol{X}$ is a vector of control variables (age and gender of respondent, predicted household income, dummies for the ways the household was sampled, and the number of social activities household members participated in during the previous three months); h represents the household, i the village, j the subdistrict, and k the audit stratum; and γ_k is a stratum fixed effect. The coefficient of interest is γ_4, which represents the differential probability in audited villages relative to control villages that family members of village government officials 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 8. Column 1 shows that family members of government officials are eight percentage points more likely to work on the project in audited villages than in control villages, and column 2 shows that family members of the project head are 14 percentage points more likely to work on the project in audited villages than in nonaudited villages. Given that the mean probability of working on the project is only 30 percent, these effects are quite large in magnitude.[28]

As discussed above, there are different interpretations for these results: one view says that this is an alternative, less desirable form of

[27] Regressions using a Probit specification produce essentially similar results.

[28] In results not reported, I find that there were no statistically significant changes in family members working on the project associated with the invitations treatments. For the comment forms treatment, there was a statistically significant increase in family members working on the project, but only for family members of the project head.

TABLE 8
NEPOTISM

	(1)	(2)	(3)	(4)
Audit	−.011	.004	−.017	−.038
	(.023)	(.021)	(.032)	(.032)
Village government family member	−.020	.016	.016	−.014
	(.024)	(.017)	(.017)	(.023)
Project head family member	.051	−.015	.051	−.004
	(.032)	(.047)	(.032)	(.047)
Social activities	.017***	.017***	.013*	.014**
	(.006)	(.006)	(.006)	(.006)
Audit × village government family member	.079**			.064*
	(.034)			(.034)
Audit × project head family member		.138**		.115*
		(.060)		(.061)
Audit × social activities			.010	.008
			(.008)	(.008)
Stratum fixed effects	Yes	Yes	Yes	Yes
Observations	3,386	3,386	3,386	3,386
R^2	.26	.26	.26	.27
Mean dependent variable	.30	.30	.30	.30

NOTE.—Data are taken from the household survey. Each observation represents one household. Results come from estimating eq. (3), where the 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 are 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 plus 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 percent.
** Significant at 5 percent.
*** Significant at 1 percent.

corruption, whereas the other suggests that 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, forthcoming); so if reducing moral hazard was the issue, one might expect effects for workers with many social connections similar to those for family members. In column 3 of table 8, however, I find that while workers with many social connections are more likely to work on the project overall, there is no statistically significant differential effect in response to the audits in the relationship between social connections and working on the project. Column 4 shows that family member results are still present when I examine all the interactions jointly. Furthermore, in results not reported here, I find that, conditional on observables, family members of village officials are more likely to be employed in the higher wage category (skilled labor rather than unskilled), suggesting that they may be receiving rents from the project. While this evidence is suggestive of a

TABLE 9
PARTICIPATION: FIRST STAGE

	Attendance (1)	Attendance of Nonelite (2)	Number Who Talk (3)	Number Nonelite Who Talk (4)
Invitations	14.83***	13.47***	.743***	.286***
	(1.35)	(1.25)	(.188)	(.079)
Invitations plus comments	11.48***	10.28***	.498***	.221***
	(1.35)	(1.27)	(.167)	(.069)
Meeting 2	−5.32***	−4.00***	.163	.024
	(1.11)	(1.06)	(.155)	(.084)
Meeting 3	−4.29***	−5.78***	.431**	−.158*
	(1.20)	(1.13)	(.172)	(.089)
Stratum fixed effects	Yes	Yes	Yes	Yes
Observations	1,775	1,775	1,775	1,775
R^2	.39	.38	.47	.28
Mean dependent variable	47.99	24.15	8.02	.94
p-value invitations = invitations + comment forms	.03	.03	.21	.43

NOTE.—Results come from estimating eq. (1), with the dependent variables the participation variables shown in the first row. Data are taken from the meeting survey. Each observation is a single village meeting. Stratum (subdistrict) fixed effects are included; since audit is constant within a subdistrict, the audit variable is automatically captured by the stratum fixed effect. Robust standard errors are in parentheses, adjusted for clustering at the village level.
* Significant at 10 percent.
** Significant at 5 percent.
*** Significant at 1 percent.

nepotism-as-corruption story, it is by no means definitive, and understanding this phenomenon is an important direction for future work.

C. *The Participation Experiments*

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—that is, the invitations and comment forms—did, in fact, increase villager participation. To examine this, in table 9 I reestimate equation (1), where the dependent variables are several measures of participation in the accountability process. Each observation represents one accountability meeting, so there are three observations for each village. Standard errors are adjusted to take into account this intravillage correlation.

Column 1 shows that the treatments had a substantial effect on total attendance: the invitations treatment increased attendance at the meetings by an average of 14.8 people, or approximately 40 percent.[29] The

[29] Villages receiving 500 invitations had slightly higher attendance (by 1.7 people on average) than villages receiving 300, though these differences are not statistically significant. Passing out invitations through schools did not result in a statistically significant difference in the composition of attendees.

slightly smaller increase for villages receiving comment forms as well as invitations suggests that being able to submit written comments and attending meetings are substitutes. Column 2 shows that virtually all of the increase in attendance at the meetings came in the form of increased attendance by these nonelite villagers, so that the number of nonelite at the meetings increased by 75 percent, from 16 people in the control villages to 29 people in the treatment villages.[30]

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 0.74, or just under 10 percent. Column 4 shows that about 40 percent of these new speakers were nonelite, increasing the number of nonelite villagers who spoke at a meeting by about 30 percent over the level in control villages.

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. 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), 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.

2. Effect on Meetings

Table 10 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 or issue that was discussed at the meeting and coded whether the problem was potentially corruption-related or not.[31] 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 to an internal village investigation, asking for help from district project officials, or requesting an

[30] I classify people as "nonelite" if they have no official position in the village, have no official position on the project, and were not described as a *tokoh masyarakat* ("informal village leader") by village members who assisted the enumerator.

[31] 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.

TABLE 10
PARTICIPATION: IMPACT ON MEETINGS

	Number of Problems (1)	Any Corruption-Related Problem (2)	Serious Response Taken (3)
Invitations	.072	.027**	−.003
	(.063)	(.013)	(.008)
Invitations plus comments	.104	.026**	.015**
	(.064)	(.012)	(.008)
Meeting 2	−.187***	.002	−.020**
	(.066)	(.013)	(.009)
Meeting 3	−.428***	−.036***	−.029***
	(.074)	(.012)	(.009)
Stratum fixed effects	Yes	Yes	Yes
Observations	1,783	1,783	1,783
R^2	.50	.31	.22
Mean dependent variable	1.18	.07	.03
p-value invitations = invitations + comment forms	.60	.96	.02

NOTE.—Results come from estimating eq. (1), with the dependent variables the outcome of meetings shown in the first row. Data are taken from the 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 to an internal village investigation, asking for help from district project officials, or requesting an external audit. Estimation is by OLS. Robust standard errors are in parentheses, adjusted for clustering by village.
* Significant at 10 percent.
** Significant at 5 percent.
*** Significant at 1 percent.

external audit. These serious actions are quite rare—they occur at only 3 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 plus comment forms treatment had a significant effect on the total number of problems discussed at the meeting. This implies that the increase in the number of people talking in table 9 is an increase in the number of people who participate in the discussion, not an increase in the number of problems per se. However, as shown in column 2, both the invitations and the invitations plus comment forms increased the probability of having a corruption-related problem discussed at the meeting by 2.7 percentage points, or 50 percent above the level in control villages. Only the comment forms, however, affected how problems were resolved: column 3 shows that the probability of a serious action being taken is 1.5 percentage points higher—or 70 percent higher than the level in control villages—in villages receiving the comment forms, but that there is no effect in villages receiving only invitations.

These results suggest that the impact of the comment forms was slightly different from what was expected. In particular, despite the large number of comments received, adding the comment forms did not change the probability that a corruption-related problem was discussed. This does not mean that fear of retaliation was not an issue, however. Rather, the results suggest that when an issue was brought up, possibly by an elite member of the village who had less fear of retaliation by the implementation team, villagers may have been unsure whether to side with the challenger or the implementation team. 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. This suggests that the comment forms were more about creating common knowledge about a problem than bringing previously unknown problems to light. Nevertheless, both of these effects were small in absolute magnitude.

3. Effect on Missing Expenditures

Table 11 examines the overall impact of the two participation treatments on the percent missing in the projects. Panel A shows the effect of the invitations treatment; panel B shows the effect of the invitations plus comment forms treatment. The results suggest that both the invitations and invitations plus comment forms treatments had a small, and statistically insignificant, impact on the overall percent missing from the project. Depending on the specification and the measure of corruption, the point estimates suggest that these treatments reduced the percent missing by between 1.5 and three percentage points, though these estimates are never statistically distinguishable from zero.

The lack of a strong effect of the invitations and comment forms on corruption is consistent with the evidence in Section V.*C*.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 anticorruption action at a meeting eliminated corruption entirely, the reduction in corruption caused by these induced serious actions would have been less than one percentage point of average total expenditures.[32] The results here suggest that not only were the direct effects (via anticorruption actions) small, but any deterrent effects of the treatments were small as well.

However, the small effects on the overall percent missing variable mask the fact that there were substantial reductions in missing labor

[32] To see this, suppose that taking a serious anticorruption action eliminates corruption in the village. (In fact, in the cross section, villages in which a serious response was taken had missing expenditures about 18 percentage points lower than other villages.) If I reestimate table 10 at the village level rather than the meeting level, the point estimate is that the comment forms treatment increased the probability of a serious response to a problem by 3.8 percentage points. This suggests that, if the comment forms had an effect only through the probability of a serious response being taken, the expected average reduction in the percent missing would be $18 \times 0.038 = 0.68$ percentage points.

expenditures. As shown in table 11, the invitations treatment led to a statistically significant reduction in missing expenditures in labor of between 14 and 22 percentage points. The invitations plus comment forms treatment also led to a reduction in missing labor expenditures, though it was somewhat smaller in magnitude (nine to 13 percentage points) and not statistically significant.[33] On the other hand, the point estimates suggest no change, or if anything a very mild increase, in missing funds on the materials dimension.[34] Since, as shown in table 3, materials account for 68 percent of road expenditures whereas unskilled labor accounts for only 20 percent, the lack of an effect on the materials dimension is why the effects on the overall percent missing are small and statistically insignificant despite the reduction in missing labor.

There are several different potential explanations for the reduction in labor expenditures but not materials associated with the participation treatments. First, it may be easier for villagers to observe actual wage payments than the quantity of materials delivered, making corruption in labor technologically easier for villagers to detect. Alternatively, even if villagers had equal information about both types of corruption, the focus on labor may have arisen 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.[35]

Though these hypotheses are difficult to definitively distinguish, there is suggestive evidence in favor of the hypothesis that it was more about incentives than information. First, the household survey described in Appendix A included a question on the respondent's perceptions of corruption in the road project. In related work (Olken 2006*b*), I find that these perceptions of corruption in the project were positively cor-

[33] I cannot reject the hypothesis that the two participation treatments (invitations and invitations plus comment forms) had the same effect. Pooling them together yields an average reduction in labor expenditures of between 12 and 17 percentage points, with the average reduction statistically significant in two of three specifications.

[34] In results not reported in the table, I find that the substitution from missing labor to missing materials is particularly pronounced in audit villages, though the interaction effects are typically not statistically significant. Using median regressions rather than OLS regressions produces somewhat smaller, and not statistically significant, changes in missing labor in response to invitations, though the qualitative finding that the response on the labor dimension is larger than the response on the materials dimension still holds.

[35] To see why reducing theft of wages might have been in the private interest of the workers, note that the typical way in which wages were stolen was to convince workers to work on the project on a volunteer basis as *swadaya masyarakat*, or community self-help, but then to bill the project as though the workers had been paid. In such a situation, if ex post this scheme was discovered, a natural form of restitution would have been to return the wages to those workers who had worked for free on the project. If so, this would give those workers who had donated some labor to the project a strong incentive to make sure that the project had not secretly been billing the project for the hours they had worked for free.

TABLE 11
PARTICIPATION: MAIN THEFT RESULTS

			No Fixed Effects		Engineer Fixed Effects		Stratum Fixed Effects	
Percent Missing[a]	Control Mean (1)	Treatment Mean (2)	Treatment Effect (3)	p-Value (4)	Treatment Effect (5)	p-Value (6)	Treatment Effect (7)	p-Value (8)
			A. Invitations					
Major items in roads (N = 477)	.252	.230	−.021	.556	−.030	.385	−.026	.448
	(.033)	(.033)	(.035)		(.034)		(.034)	
Major items in roads and ancillary projects (N = 538)	.268	.236	−.030	.360	−.032	.319	−.029	.356
	(.031)	(.031)	(.032)		(.032)		(.032)	
Breakdown of roads:								
Materials (N = 477)	.209	.221	.014	.725	.008	.839	.005	.882
	(.041)	(.041)	(.038)		(.037)		(.037)	
Unskilled labor (N = 426)	.369	.180	−.187*	.058	−.215**	.024	−.143*	.098
	(.077)	(.077)	(.098)		(.094)		(.086)	

	B. Invitations Plus Comments							
Major items in roads (N = 477)	.252	.228	−.022	.455	−.024	.411	−.015	.601
	(.033)	(.026)	(.030)		(.029)		(.030)	
Major items in roads and ancillary projects	.268	.238	−.026	.409	−.025	.406	−.027	.385
(N = 538)	(.031)	(.026)	(.032)		(.030)		(.031)	
Breakdown of roads:								
Materials (N = 477)	.209	.180	−.028	.414	−.022	.496	−.010	.754
	(.041)	(.032)	(.034)		(.032)		(.033)	
Unskilled labor (N = 426)	.369	.267	−.099	.255	−.132	.131	−.090	.323
	(.077)	(.073)	(.087)		(.087)		(.091)	

NOTE.—See the note to table 4. Results come from estimating eq. (1), a regression of the dependent variable on a dummy for audit treatment, invitations treatment, and invitations plus comment forms treatments. Each invitations effect and invitations plus comments effect comes from a separate regression, with the dependent variable listed in the row and the fixed effects specification listed in the column heading. Robust standard errors are in parentheses. Regressions without stratum (i.e., subdistrict) fixed effects include a variable for audits and allow for clustering of standard errors by subdistrict.

[a] Percent missing equals log reported value − log actual value.

* Significant at 10 percent.

** Significant at 5 percent.

*** Significant at 1 percent.

related with price markups in materials expenditures but were essentially uncorrelated with missing labor expenditures or with quantities of materials. This suggests that, if anything, villagers appear to have more information about missing materials, not missing labor expenditures.

Second, one can examine whether the invitations treatment varies when workers come from outside the village and are therefore ineligible to attend the accountability meetings. When I do this (i.e., interact the invitations treatment with the percentage of workers coming from outside the village), I find that the reduction in missing labor associated with the invitations treatment is smaller when more of the workers come from outside the village. Since the only people who attend the accountability meetings are those who live in the village, this suggests that it is the fact that the invitations induce workers to attend meetings and fight for their own wages that drives the reduction in theft in wages. The percentage of workers coming from outside the village is potentially endogenous, so these results should be viewed as speculative. Nevertheless, this result, combined with the result about information, suggests that the lack of a reduction in materials may be caused by the fact that monitoring theft of materials is more of a public good than monitoring theft of wages.

4. Elite Capture and the Participation Treatments

To the extent that the village elites are capturing the invitations process to prevent effective monitoring, one would expect that to whom the invitations or comment forms were distributed might affect their effectiveness. As described above, invitations and comment forms were distributed in one of two ways: either through primary schools in the villages or via the neighborhood heads affiliated with the village government. Since the distribution method used was also randomized between villages, I can investigate whether distributing the invitations and comment forms via the schools—which reaches an effectively random sample of village members—was more effective than distributing the invitations and comment forms via the neighborhood heads, which may have allowed the village elite to retain their capture of the program even while appearing to be inclusive.

In table 12, I investigate the impact of the invitations and invitations plus comment forms treatments separately by the method through which they were distributed. The invitations treatment shows no statistically significant impact in either case. The comment form treatment, however, appears to lead to a statistically significant reduction in missing expenditures when forms were distributed via schools (treatment effects between -0.052 and -0.086, depending on the specification), whereas it shows no impact when they were distributed via neighborhood heads.[36]

These results suggest that, when given the opportunity, neighborhood heads channeled comment forms to those villagers predisposed to be favorable to the project, thus diminishing their effect. This type of elite capture can be examined more directly by examining the contents of the forms that were returned. In particular, as discussed above, the comment forms included three closed-ended questions, asking the respondent to rate on a scale of 1 to 3 the prices paid by the project, the quality of the financial management of the project, and the quality of the construction on the project. I take the average scores on the three questions and normalize them to range from 0 (worst) to 1 (best). The average score among all returned comment forms was 10 percent lower when the forms were distributed via schools than when they were distributed by neighborhood heads (0.39 and 0.43, respectively; p-value <0.001). Thus, when comment forms were distributed via neighborhood heads, the comments received were more positive, even though missing expenditures were actually *higher* in these villages. These results suggest that elite capture of the monitoring process may be an important reason why it is not always more effective.

Even without the invitations and comment forms, the KDP program investigated here includes more grassroots 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—that is, the effects reported above—do not capture the overall effects of grassroots participation as a monitoring mechanism, and that more dramatic variations in the amount of participation might have different effects. However, the fact that the invitations did lead to a substantial reduction in missing labor expenditures and that comment forms did lead to a substantial reduction in missing expenditures when distributed via schools suggests that the changes in participation induced by the interventions were large enough to make a difference. More generally, the results suggest that for grassroots monitoring to succeed, care must be taken to reduce the free-rider problem and limit the ability of local elites to capture the process.

[36] In results not reported in the table, I find that the difference in the comment form treatment effect between cases in which it was distributed via neighborhood heads and cases in which it was distributed via schools is generally statistically significant at conventional levels (p-values between 0.006 and 0.119, depending on the specification).

TABLE 12
INTERACTIONS OF PARTICIPATION EXPERIMENTS WITH HOW INVITATIONS WERE DISTRIBUTED

PERCENT MISSING[a]	CONTROL MEAN (1)	TREATMENT MEAN (2)	NO FIXED EFFECTS: Treatment Effect (3)	NO FIXED EFFECTS: p-Value (4)	ENGINEER FIXED EFFECTS: Treatment Effect (5)	ENGINEER FIXED EFFECTS: p-Value (6)	STRATUM FIXED EFFECTS: Treatment Effect (7)	STRATUM FIXED EFFECTS: p-Value (8)
	A. Invitations							
	Invitations Distributed via Neighborhood Heads							
Major items in roads (N = 246)	.252	.222	−.030	.469	−.043	.274	−.042	.324
	(.033)	(.044)	(.042)		(.039)		(.043)	
Major items in roads and ancillary projects (N = 271)	.268	.255	−.013	.761	−.015	.712	−.004	.924
	(.031)	(.045)	(.043)		(.041)		(.043)	
	Invitations Distributed via Schools							
Major items in roads (N = 233)	.252	.239	−.009	.854	−.014	.774	−.003	.950
	(.033)	(.046)	(.050)		(.048)		(.045)	
Major items in roads and ancillary projects (N = 263)	.268	.216	−.048	.282	−.051	.245	−.056	.155
	(.031)	(.040)	(.044)		(.043)		(.039)	

	B. Invitations Plus Comments							
	Invitations Plus Comment Forms Distributed via Neighborhood Heads							
Major items in roads (N = 242)	.252	.278	.025	.483	.038	.294	.022	.602
	(.033)	(.036)	(.036)		(.036)		(.041)	
Major items in roads and ancillary projects (N = 271)	.268	.277	.010	.792	.024	.535	.023	.569
	(.031)	(.039)	(.039)		(.038)		(.040)	
	Invitations Plus Comment Forms Distributed via Schools							
Major items in roads (N = 242)	.252	.179	−.070*	.093	−.086**	.023	−.052	.150
	(.033)	(.036)	(.041)		(.038)		(.036)	
Major items in roads and ancillary projects (N = 267)	.268	.198	−.064	.127	−.077*	.052	−.078*	.056
	(.031)	(.034)	(.042)		(.039)		(.041)	

NOTE.—See the note to table 11. N refers to the number of observations with nonmissing dependent variable for control observations (i.e., where neither invitations nor invitations and comment forms were distributed) plus the number of treatment observations for the listed treatment. Treatment effects and p-values are computed from a regression of the dependent variable on a dummy for audit treatment and four dummies for the participation treatments interacted with distribution mechanism (i.e., invitations distributed via neighborhood heads, invitations distributed via schools, invitations plus comment forms distributed via neighborhood heads, and invitations plus comment forms distributed via schools).

[a] Percent missing equals log reported value − log actual value.

* Significant at 10 percent.

** Significant at 5 percent.

*** Significant at 1 percent.

TABLE 13
COST-BENEFIT ANALYSIS OF AUDIT TREATMENT

	Equal-Weighted Net Benefits (1)	Distribution-Weighted Net Benefits (2)
Cost of treatment:		
Monetary cost	−335	−278
Associated deadweight loss	−134	−111
Time cost	−31	−31
Subtotal	−500	−419
Change in rents received by corrupt officials:		
From theft of materials	−367	−224
From theft of wages	−102	−62
Subtotal	−468	−286
Change in benefits from project:		
Net present value of road expenditures	1,165	1,165
Wages received by workers	86	86
Other expenditures	−37	−37
Subtotal	1,213	1,238
Total net benefits	245	508

NOTE.—All figures are given in U.S. dollars. Costs are listed as negative numbers. Distributional weights were calculated using CRRA utility with a coefficient of relative risk aversion of two. Derivation of change in the net present value of the road and additional assumptions are discussed in the text.

VI. Cost-Benefit Analysis

The previous analysis discussed the direct effects of the treatments. 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 missing expenditures, and therefore should be viewed as somewhat more speculative than the preceding sections. I focus on the audit treatment since that was the treatment with the strongest average effects.

Table 13 presents the cost-benefit estimates. I present two sets of net benefits: "equal-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 whereas others are borne by the poor.[37]

[37] Specifically, the distribution-weighted net benefits assume constant relative risk aversion (CRRA) utility of per capita consumption with a coefficient of relative risk aversion of two, normalized so that the median household in rural Java has marginal utility of one. Households in which someone is in the village government have per capita expenditures 18.5 percent higher than typical households in the village, so this (plus the CRRA assumptions) suggests that the social value of $1.00 of rents received by project or government officials is $0.61. On the other hand, the social value of $1.00 of increased wages

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. The associated deadweight loss is the deadweight loss associated with the increased taxes required to pay the monetary cost of the treatments.[38] Finally, the time cost is the monetary value of the additional time villagers spend at village meetings as a result of the audit treatments, valued at average local wage rates.

The estimates for the change in rents received by corrupt officials are taken from column 3 in table 4, multiplied by the average cost of the project and the average percentage of the project consisting of materials and labor expenditures from table 3. In partial equilibrium, the social value of these rents is one since they can be consumed by village officials. However, qualitative evidence suggests that much of the rents from corruption is dissipated ex ante during the campaign for village head, so in general equilibrium a reduction in the rents from corruption would translate into a reduction in village head campaign expenditures. The key question in assigning a social value to the change in corruption then becomes the form of campaign expenditures: if they take the form of cash handouts to villagers, the social value of campaign expenditures (and hence of rents from corruption) would be one; if they take the form of posters with the candidates' name on them, the social value of campaign expenditures (and hence of rents from corruption) would be close to zero. To be conservative, for the moment I assume that the social value is one; assuming a social value of zero would make the audits appear more cost-effective.

To estimate the change in benefits from the project, I use the point estimates of the change in actual expenditures relative to the original (prerandomization) plan.[39] For wages, the increase in actual wage expenditures is valued at the actual amount transferred. For materials, the

received by workers is $1.29, since 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 one. Finally, for taxes, I estimate the distributional impact of both the monetary and deadweight loss cost using national consumption data from the 2003 Survei Sosial Ekonomi Nasional (SUSENAS), assuming that the burden is borne proportionally to consumption.

[38] Estimates of the marginal cost of public funds for indirect taxes vary considerably. For developing countries, estimates for the marginal cost of funds from indirect taxes range from 1.04–1.05 in Indonesia and Bangladesh (Devarajan, Thierfelder, and Suthiwart-Narueput 2002) to 1.59–2.15 for India (Ahmad and Stern 1987). By comparison, estimates of the marginal cost of public funds for the United States range from 1.17 to 1.56 (Ballard, Shoven, and Whalley 1985), with policy analysis typically using values in the range from 1.30 to 1.40. I assume that the marginal cost of public funds is 1.4 (i.e., the deadweight loss is 0.4)

[39] The reason to use the change in actual expenditures relative to the original plan, rather than missing expenditures in each category, is that there may have been changes in which corruption was hidden in the accounting ledgers, and such nominal changes in reporting would have no efficiency consequences.

increase in materials expenditure increases the life span of the road.[40] To value the change in life span 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 US$3.41.[41] For other reported expenditures, I assume a marginal social value of one since it is unclear what percentage of them are actually spent productively on the road as opposed to being captured by local officials.

The results presented in table 13 suggest that the audits were substantially cost-effective. I estimate that the *net* social benefits from the audits were approximately $250 per village, which implies that the benefits were more than 150 percent of the cost of the audits. Weighting by distributional incidence increases these net benefits even more.

An alternative method of calculating the cost-benefit analysis that relies on fewer assumptions (and that in many ways better approximates the way government agencies actually make such decisions) would be simply to compare the reduction in corruption to the cost of the audits. This implicitly assumes that the social value of transfers to corrupt village officials is zero. It also assumes that, rather than have the road be degraded from being constructed at lower quality, the government would simply increase the project budget to make up for anticipated corruption.[42] Under such a scenario, the relevant comparison is the reduction in corruption due to the audits—$468 per village—as compared to the cost of the audits—$335 per village (or $366 if we include villagers' time costs). With this simpler methodology, the audits once again appear cost-effective.

The audit treatment discussed here was a move from a 4 percent baseline audit probability to a 100 percent audit probability. It is pos-

[40] A natural question is why, in equilibrium, the government does not simply increase the road budget to compensate for the average level of corruption and ensure that the equilibrium road is built efficiently. One reason is that the percentage of the budget lost to corruption may increase as the budget increases, making such a policy not feasible. In fact, from the data, this appears to be the case: I find that the percent missing is higher when the ex ante budget per kilometer is higher, either using OLS or instrumenting for the budget per kilometer with the number of villages in the subdistrict competing over the same grant amount per subdistrict. Chavis (2006) also finds similar effects looking at unit costs on a wider range of KDP projects.

[41] Dent estimates that KDP road projects produce annual flow benefits of 33 percent 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 five years. On the basis of conversations with KDP engineers, I assume that each percentage of materials stolen reduces the life of the road by 0.1 year, so stealing 50 percent of materials is the difference between good and poor quality. Assuming a real discount rate of 5 percent and a baseline life span of a road of seven years implies that each dollar of materials stolen reduces the discounted net benefits of the project by $3.41.

[42] As a result, we do not need an assumption for the marginal cost of public funds, since that assumption would apply equally to the additional revenue required to make up for corrupt losses or to pay for the audits.

sible, however, that the response of corruption to the audit probability is concave, so that an audit probability of only 50 percent or even 25 percent might achieve most of the benefits.[43] The costs of audits, however, are roughly linear in the audit probability. This suggests that raising audit probabilities to an intermediate level, rather than all the way to 100 percent, might have an even higher cost-benefit ratio than the 100 percent audit probability documented here.

VII. 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 grassroots participation in the village monitoring process.

The evidence suggests that increasing the probability of external audits substantially reduced missing funds in the project. In particular, increasing the probability that a village was audited by the central government audit agency from a baseline of 4 percent to 100 percent reduced missing expenditures from 27.7 percentage points to 19.2 percentage points. One reason that the decrease was not larger is that a 100 percent audit probability does not imply that village officials face a 100 percent probability of detecting corruption and imposing a punishment. In fact, although auditors found violations of some type or another in 90 percent of the villages they visited, the vast majority of these violations were procedural in nature, and there were very few, if any, cases in which the auditors had enough concrete evidence to actually prosecute corrupt offenses. The low probability of a formal prosecution and punishment suggests that higher punishments conditional on prosecution may be an effective complement to higher audit probabilities. They also suggest that providing audit results to the public, who can then use them in making their electoral choices, may be a useful complement to formal punishments.

The evidence on grassroots participation showed that increasing grassroots participation in monitoring reduced missing expenditures only under a limited set of circumstances. First, the results showed that inviting more villagers to monitoring meetings reduced only missing labor expenditures, with no impact on materials and, as a consequence, little impact overall. Since a small group of laborers stands to gain from reducing corruption in labor, whereas the entire village stands to gain from reducing corruption in materials, this suggests that grassroots mon-

[43] 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.

itoring can be effective in circumstances in which there is relatively little free-riding. For example, 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, may be appropriate candidates for grassroots monitoring. For public goods in which incentives to monitor are much weaker, such as the infrastructure projects studied here, the results suggest that using professional auditors may be much more effective.

Second, the results showed that issuing anonymous comment forms to villagers reduced missing expenditures only if the comment forms were distributed via schools in the village, completely bypassing village officials who may have been involved in the project. This suggests that care must be taken in designing grassroots monitoring programs to ensure that they are not captured by local elites.

The results in this paper represent the results from a short-run intervention. If auditors are bribable, over time villages may develop repeat relationships with auditors that 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.

Even for this one-time intervention, certain results will become apparent only with time. For example, after several years, it will be apparent whether the increased scrutiny imposed by the audits affects who chooses to become involved in project management, and whether negative audit findings affect the reelection probabilities of village officials. Reducing corruption may also reduce campaign expenditures for village offices, since the rents from obtaining these positions will have declined. Whether the reduced campaign expenditures take the form of fewer cash handouts to villagers, or fewer banners advertising the candidates' names, will determine the ultimate general equilibrium social welfare implications of the reduction in corruption. The efficiency impact of the reduction in corruption will also become clearer with time since we can observe changes in how long the road lasts. Understanding the long-run implications of anticorruption policies remains an important issue for future research.

Appendix A

Data Collection

In addition to the corruption field survey described in Section IV 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 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 begun plus the three accountability meetings. At each meeting, the enumerator circulated an attendance list. 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 designation typically given to teachers, religious leaders, and other types of informal village leaders.

While the meeting was in progress, the enumerator was asked to keep detailed notes on what occurred during the meeting. The enumerator compiled 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 seven 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 predefined 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. Household expenditures were predicted on the basis of assets, using the relationship between assets and consumption from the 1999 Survei Seratus Desa (Hundred Villages Survey). The household survey was designed as a stratified random sample, containing between six and 13 respondents per village.

Field Survey/Corruption Measurement

The general approach used in the field survey is described in Section IV of the text. 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 percent of total project costs. Of these voluntary contributions, the bulk (60 percent 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, since 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 IV, for labor, I use only reports and actual estimates of paid labor and exclude voluntary labor.

For the worker survey described in Section IV, a total of five worker interviews were conducted in each village. Of these interviews, two were focus group interviews, consisting of three or more workers interviewed together, and three were individual interviews. Two of these interviews (one focus group, one individual) involved workers randomly selected from the official list of all workers who had ever been paid by the project; the remainder were recruited more informally, by having the surveyor go to different areas of the village located near the road and asking household members 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 worked only 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 appendix describes the assumptions used in the paper in more detail 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 and 23 percent for 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

TABLE B1
ASSUMPTIONS FOR LOSS RATIOS AND WORKER CAPACITY

	Results from Calibration
Loss ratios:	
Sand	1.00
Rocks	1.20
Gravel	1.75
Worker capacity:	
Clearing brush and cleaning road surface (m^2)	20
Spreading sand (m^3)	4.5
Splitting rocks (m^3)	3.0
Installing rock layer (m^2)	6.5
Spreading gravel (m^3)	2.25
Digging side channels and creating shoulder (m^3)	1.0
Building retaining wall (m^3) (unskilled labor portion)	.33*
Standard cut and fill (m^3)	2*

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

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 four 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 B1 lists the main assumptions used in the study, based on the results of the calibration. For the labor estimates, the revised estimates were revised downward 20 percent 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 suggest 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 plus gravel" together (with an implied joint loss ratio of 1.25) rather than separately. Doing so reduces the average percent missing from 24 percent to 22 percent, but otherwise does not alter the results of the paper.

References

Ahmad, Etisham, and Nicholas Stern. 1987. "Alternative Sources of Government Revenue: Illustrations from India, 1979–80." In *The Theory of Taxation for Developing Countries*, edited by David Newbery and Nicholas Stern. New York: Oxford Univ. Press (for World Bank).

Ballard, Charles L., John B. Shoven, and John Whalley. 1985. "General Equilibrium Computations of the Marginal Welfare Costs of Taxes in the United States." *A.E.R.* 75 (March): 128–38.

Bardhan, Pranab. 2002. "Decentralization of Governance and Development." *J. Econ. Perspectives* 16 (Fall): 185–205.

Bardhan, Pranab, and Dilip Mookherjee. 2006. "Decentralisation and Accountability in Infrastructure Delivery in Developing Countries." *Econ. J.* 116 (January): 101–27.

Becker, Gary S., and George J. Stigler. 1974. "Law Enforcement, Malfeasance, and Compensation of Enforcers." *J. Legal Studies* 3 (January): 1–18.

Cadot, Olivier. 1987. "Corruption as a Gamble." *J. Public Econ.* 33 (July): 223–44.

Chavis, Larry. 2006. "Decentralizing Development: Allocating Public Goods via Competition." Manuscript, Univ. North Carolina, Chapel Hill.

Dent, Geoffrey. 2001. "Ex-Post Evaluation of Kecamatan Development Program (KDP) Infrastructure Projects." Manuscript, Project Appraisals Pty, Jakarta.

Devarajan, Shantayanan, Karen E. Thierfelder, and Sethaput Suthiwart-Narueput. 2002. "The Marginal Cost of Public Funds in Developing Countries." In *Policy Evaluation with Computable General Equilibrium Models*, edited by Amedo Fossati and Wolfgang Wiegard. New York: Routledge.

Di Tella, Rafael, and Ernesto Schargrodsky. 2003. "The Role of Wages and Auditing during a Crackdown on Corruption in the City of Buenos Aires." *J. Law and Econ.* 46 (April): 269–92.

Fisman, Raymond, and Shang-Jin Wei. 2004. "Tax Rates and Tax Evasion: Evidence from 'Missing Imports' in China." *J.P.E.* 112 (April): 471–96.

Hsieh, Chang-Tai, and Enrico Moretti. 2006. "Did Iraq Cheat the United Nations? Underpricing, Bribes, and the Oil for Food Program." *Q.J.E.* 121 (November): 1211–48.

Karlan, Dean. Forthcoming. "Social Capital and Group Banking." *Econ. J.*

Krueger, Anne O. 1974. "The Political Economy of the Rent-Seeking Society." *A.E.R.* 64 (June): 291–303.

Lambsdorff, Johann G. 2003. "Background Paper to the 2003 Corruption Perceptions Index." Manuscript, Transparency Internat., Passau, Ger.

Mauro, Paolo. 1995. "Corruption and Growth." *Q.J.E.* 110 (August): 681–712.

Nagin, Daniel S., James B. Rebitzer, Seth Sanders, and Lowell J. Taylor. 2002. "Monitoring, Motivation, and Management: The Determinants of Opportunistic Behavior in a Field Experiment." *A.E.R.* 92 (September): 850–73.

Olken, Benjamin A. 2005. "Monitoring Corruption: Evidence from a Field Experiment in Indonesia." Working Paper no. 11753 (November), NBER, Cambridge, MA.

———. 2006*a*. "Corruption and the Costs of Redistribution: Micro Evidence from Indonesia." *J. Public Econ.* 90 (May): 853–70.

———. 2006*b*. "Corruption Perceptions vs. Corruption Reality." Working Paper no. 12428 (July), NBER, Cambridge, MA.

Reinikka, Ritva, and Jakob Svensson. 2004. "Local Capture: Evidence from a Central Government Transfer Program in Uganda." *Q.J.E.* 119 (May): 679–705.

———. 2005. "Fighting Corruption to Improve Schooling: Evidence from a Newspaper Campaign in Uganda." *J. European Econ. Assoc.* 3 (April–May): 259–67.

Rose-Ackerman, Susan. 2004. "The Challenge of Poor Governance and Corruption." Paper presented at the Copenhagen Consensus Challenge, May.

Shleifer, Andrei, and Robert W. Vishny. 1993. "Corruption." *Q.J.E.* 108 (August): 599–617.

Stiglitz, Joseph E. 2002. "Participation and Development: Perspectives from the Comprehensive Development Paradigm." *Rev. Development Econ.* 6 (June): 163–82.

Woodhouse, Andrea. 2004. "Village Justice in Indonesia: Case Studies on Access to Justice, Village Democracy and Governance." Manuscript, World Bank, Washington, DC.

World Bank. 2004. *Making Services Work for Poor People.* World Development Report 2004. Washington, DC: World Bank.

Yang, Dean. 2004. "Can Enforcement Backfire? Crime Displacement in the Context of Customs Reform in the Philippines." Manuscript, Univ. Michigan.