Education and Health Care in Developing Countries

by

Trang V. Nguyen

B.A. Economics and Mathematics
Brandeis University, 2003

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2008

Signature of Author.................................
Department of Economics
August 1, 2008

Certified by.................................
Esther Duflo
Abdul Latif Jameel Professor of Poverty Alleviation and Development Economics
Thesis Supervisor

Certified by.................................
Abhijit Banerjee
Ford International Professor of Economics
Thesis Supervisor

Accepted by.................................
Peter Temin
Elisha Gray II Professor of Economics
Chairman, Departmental Committee on Graduate Studies

Education and Health Care in Developing Countries

by

Trang V. Nguyen

Submitted to the Department of Economics
on August 15, 2008, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This thesis is a collection of three essays on education and health in developing countries.

Chapter 1 shows that increasing perceived returns to education strengthens incentives for schooling when agents underestimate the actual returns. I conducted a field experiment in Madagascar to study alternative ways to provide additional information about the returns to education. I randomly assigned schools to the role model intervention, the statistics intervention, or a combination of both. I find that providing statistics reduced the large gap between perceived returns and the statistics provided. As a result, it improved average test scores and student attendance. For those whose initial perceived returns were below the statistics, test scores improved by 0.37 standard deviations. Seeing a role model of poor background has a larger impact on poor children's test scores than seeing someone of rich background. The key implication of my results is that households lack information, but are able to process new information and change their decisions in a sophisticated manner.

Chapter 2, joint work with Gerard Lassibille, evaluates several interventions in Madagascar that sought to promote top-down and local monitoring of the school to improve education quality. Randomly selected school districts and subdistricts received operational tools to facilitate their supervision tasks. Randomly selected schools in these treated districts were reinforced with teacher tools and parent-teacher meetings centered around a school report card. We find little impact of targeting district and sub-district administrators. Meanwhile, the intervention implemented at the school level improved some of the teachers' behaviors and student attendance. Student test scores also improved by 0.1 standard deviations after two years. These results suggest that beneficiary monitoring is more effective than mediated control in the hands of government bureaucrats in this context.

Chapter 3 studies informal payments to doctors and nurses for inpatient health care in Vietnam. Exploiting within-hospital variation, I find that acute patients, despite having a presumably higher benefit of treatment, are 8 percentage points less likely to pay bribes, and pay less, than non-acute patients. One plausible interpretation is that doctors might face existing incentives against neglecting acute cases. I find that the differential payment by acute status is larger in central locations (expected to be well-monitored) and at facilities that receive more audit visits. Overall, these findings may be a sign of bureaucrats responding to incentives, even in a highly corruptible environment.

Thesis Supervisor: Esther Duflo
Title: Abdul Latif Jameel Professor of Poverty Alleviation and Development Economics

Thesis Supervisor: Abhijit Banerjee
Title: Ford International Professor of Economics

Acknowledgments

I am very grateful to my advisors, Esther Duflo, Abhijit Banerjee, and Tavneet Suri. Their passion for development economics and continual support have inspired and guided me throughout my graduate studies, by much more than what can be expressed here.

This thesis has benefited from discussions with the faculty and students at MIT and other institutions. David Autor, Suman Basu, Jim Berry, Miriam Bruhn, Quoc-Anh Do, Quy-Toan Do, Rachel Glennerster, Michael Greenstone, Rick Hornbeck, Rob Jensen, Cinthya Kinnan, Ben Olken, Rob Townsend, and participants of the MIT development and applied micro lunches have provided very valuable suggestions. My research in Madagascar would not have been possible without the diligent and excellent work by the AGEMAD team at the Madagascar Ministry of Education, Jee-Peng Tan, Cornelia Jesse, and Gerard Lassibille. I am grateful to Greg Fischer and Raymond Guiteras for their comments, everyday encouragements, and ready hands of help. I thank especially Maisy Wong for sharing her kind support, insights, and friendship.

I have been fortunate to receive generous financial support from the World Bank, AFD, UNICEF, and the TOTAL Fellowship for part of the research in this dissertation.

I would like to thank all my friends and family for their optimism and understanding during the past five years. Having Tatiana Didier, Chris Smith, and Tom Wilkening as classmates and housemates made graduate school an enjoyable experience. My partner Duc, my brother Quan, and my Murray family made me laugh and kept my life balanced during tough times.

I dedicate this dissertation to my parents, Nguyen Xuan Phuc and Nguyen Thi Thuy Van. Their love and belief in me make it a worthwhile path through every challenge and success.

Contents

Chapter 1

Information, Role Models and Perceived Returns to Education: Experimental Evidence from Madagascar

1.1 Introduction

Universal primary enrollment is one of the Millennium Development Goals, and many countries, especially those in Africa, have devoted substantial efforts to attaining this objective. Nonetheless, low schooling persists even though market returns to education appear to be high and direct costs are low.[1] For example, in Madagascar, the estimated returns to one extra year in primary and secondary school are 5% and 12%, respectively. Primary education is free; yet, 40% of children entering first grade actually complete five years of primary school. Only half of those children who complete primary school continue on to secondary school. Even when children are enrolled, low learning, as reflected in the 63% pass rate at the primary-cycle examination, is an important concern (Tan 2005). In addition to low achievements, widespread student absenteeism during the school year suggests low effort on students' part.

While there can be several explanations for low schooling such as credit constraints, high discount rates, or low school quality,[2] I focus on another possibility: an information gap between perceived returns to schooling (what people consider to be the returns) and actual returns. Such a gap can exist due to costly gathering of information in isolated areas of developing countries. My paper documents this information gap in rural Madagascar. Households appear to have imperfect information about earnings associated with different levels of education, even in the presence of heterogenous returns. They will choose low education when they think the return is low (Foster and

[1] Psacharopoulos (1985) and Psacharopoulos (1994) estimate the returns to be high in many developing countries.

[2] For example, Oreopoulos (2003) explores the possible role of high discount rates in the decision of high-school dropouts. See Glewwe and Kremer (2005) for a more complete review.

Rosenzweig (1996) and Bils and Klenow (2000)). Thus, we would expect that increasing perceived returns can strengthen incentives for schooling for those who may have initially underestimated the returns.

Once we know there is imperfect information about the returns to education, the next important step is to examine how to provide useful additional information. One straightforward way is to provide statistics. The downside of this approach is that a largely illiterate population may not effectively process statistical numbers. An alternative way of informing households of the benefits of education is through a "role model," i.e. an actual person sharing his success story. This policy choice has been more popular than presenting numbers (for example, UNICEF has role model programs in various developing countries). Role models can be effective simply because stories are powerful, or because they contain information. Wilson (1987) proposes that role models bring back missing information about the upside distribution of returns to education. The impact of observing a role model will depend on how households update their beliefs based on the information the role model brings. Ray (2004) suggests that role models might be powerful only when they come from a similar background and, therefore, carry information relevant to the audience.[3]

My paper sheds further light on how to provide additional information about the returns to education. I ask the following three questions: (i) Are households' perceived returns to education different from the estimated average return, due to heterogeneity entirely or also imperfect information? (ii) How do households update beliefs when presented with statistics about the average return or with a specific role model? (iii) How do children adjust their efforts in response to the change in perceived returns? To address these questions, I first administered a survey in rural Madagascar to measure parents' perceived returns for their child and perceived returns for the average person in the population.[4] I then conducted a field experiment in 640 primary schools, in

[3]Discussions on role models have featured both in the policy environment and in the academic literature. Most of the evidence to date focuses on a "mentoring" role model; for example, a teacher of the same race has a positive impact on a student's test scores (Dee 2004). My work studies a role model's informational effect, i.e. how people update beliefs based on the information the role model brings.

[4]This survey methodology follows the literature in the US on measuring perceived returns to education. Manski

conjunction with the Madagascar Ministry of Education. The experimental design was motivated by a model of schooling decisions with heterogeneous returns and uncertainty about the return to education. When an agent is presented with the statistics, he updates his belief about the average return. When presented with a role model, he infers that heterogeneity in the returns to education is high, making the statistical estimate of the average return less relevant. Thus, role models will undermine the impact of the statistics intervention. The agent also infers from the role model's background that individuals from this background receive relatively high returns to education on average.

To test the theory's key predictions, I randomly assigned schools into three main treatment groups. In the "statistics" schools, teachers reported to parents and children the average earnings at each level of education, as well as the implied gain. The second intervention sent a role model to share with students and families his/her family background, educational experience, and current achievements. To investigate the proposition that a role model from the same background carries relevant information, I randomly assigned a role model of poor or rich background to different schools. The third treatment combines both the statistics and role model interventions together, investigating the possibility that role models may undermine statistics. I collected endline data on perceived returns, student attendance, and test scores to evaluate the impact of the interventions on beliefs and on effort at school roughly five months later.

I find that parents' median perceived return matches the average return estimated from household survey data. Nonetheless, there is a lot of dispersion in both perception of the average return and perception of the child's own return. There are two possible explanations for dispersion in perception of the child's own return: heterogeneity in the actual returns and imperfect information. I argue that while heterogeneity plays a role, some extent of imperfect information exists, as

(1992) first brought this topic to attention by proposing that youths infer the returns from their sample of observations. Several papers have surveyed college students in the US for their perceptions of incomes associated with different educational levels: Betts (1996), Dominitz and Manski (1996), Avery and Kane (2004). Most find that American students estimate quite well the mean earnings for a cohort, and think they would earn slightly better than the mean.

reflected in the larger dispersion in perception than in the real distribution of earnings.

The main results of the experiment match the model's predictions. Participants who received the statistics intervention updated their perceived returns. Providing statistics significantly decreases the gap between perceived returns and the estimated average return provided. This result holds for both perception of the average return and perception of one's own return. It suggests that part of the dispersion in perceived returns comes from imperfect information. Once households update their perceived returns, schooling decisions consequently respond. The statistics intervention improved average test scores by 0.2 standard deviations, only a few months later. For those whose initial perceived returns were below the statistics, test scores improved by 0.37 standard deviations. Student attendance in statistics schools is also 3.5 percentage points higher than attendance in schools without statistics.

The role model interventions offer very interesting results. By themselves, role models have small effects on average, but people seem to care about the information the role model brings. In particular, consistent with the theory, the role model's background matters. The role model from a poor background improved average test scores by 0.17 standard deviations, while the role model from a rich background had no impact. Moreover, this positive impact of poor-background role models on average test scores mainly reflects their influence on the poor (0.27 standard deviations). Again as the theory predicts, combining a role model with statistics undoes the effects of statistics. The role model may indicate high underlying heterogeneity in individual returns, implying that the statistics were imprecisely estimated. Thus, households do not update as much based on the statistics as they would otherwise.

These results have strong implications. First, households update their perception in a sophisticated way based on the information provided. Second, schooling investment seems responsive to changes in perceived returns. Third, in terms of policies to improve education in developing countries, providing statistical information can be a cost-effective instrument to enhance children's efforts at school, in contexts where individuals underestimate the returns. A quick back-of-theenvelope calculation using my results shows that the statistics intervention would cost 2.30 USD for an additional year of schooling and 0.04 USD for additional 0.10 standard deviations in test scores (cheaper than any prior programs evaluated with a randomized design). Lastly, when households do have correct perception of the returns on average, the results from this paper suggest that market interventions to improve the overall returns are an important way to increase school attendance and test scores.

Some existing evidence already shows that the provision of information may affect individual behaviors, such as Dupas (2006). In particular, Jensen (2007) demonstrates from a field experiment in the Dominican Republic that providing students with mean earnings by education led to a 4 percentage point increase in the probability of returning to school the following year. My work builds on this result along two fronts. First, I test whether statistics given to the parents affects the intensive margin of schooling–student attendance and test scores. Second, in terms of the research question, I study whether role models are effective in changing behaviors, either through their success story or through the relevance of their information. Addressing this research question would provide more depth to our understanding of how people update their belief.

The rest of this paper is organized as follows. Section 1.2 discusses a model of schooling decisions under Bayesian updating of perceived returns. In Section 1.3, I describe the field experiment: the statistics and role model programs, as well as the evaluation design. Section 1.4 discusses the data. Section 1.5 gives an overview of perceived returns to education at baseline, and then presents the estimation strategy and results. The final section concludes.

1.2 A Model of Schooling with Uncertainty about the Return to Education

To assess the conditions under which the statistics and role model programs may plausibly affect decisions, this paper integrates a simple framework of schooling with Bayesian updating about the return to education. This theory builds on the standard Card (1999) model of school choice with heterogeneity in individual returns to education. Here I formalize the possibility that agents measure their own return as well as the average return with errors. As a result, I explore two features of learning.

First, this section models learning about the average return to education. I show that agents update their estimate after observing the government's statistic on the average return. How much they update depends on their belief about the precision of the government's estimate. When there is high heterogeneity in individual returns in the population, the government's estimate of the average return is less precise. In that case, agents should put less weight on this statistic.

Second, an agent also learns about the relationship between his type and his return to education, above and beyond the population's average return. He infers this information from observing the government's choice of role model programs. School investment decisions are then made based on the posterior belief about the return to education.

1.2.1 Model Setup

Consider individual schooling choice under heterogeneous returns to education. An individual i has the following preferences[5]

$$EU = E_i \ln y_i(e_i) - c_i(e_i) \tag{1.1}$$

where log earnings is a linear function of education $\ln y_i(e_i) = a_i + b_i e_i$ and $c_i(e_i)$ denotes an increasing and convex cost function of education. In the standard model of investment in human capital, e_i represents years of schooling. Here I examine effort behaviors of children already enrolled in school. I refer to e_i as child effort in schooling, but the general intuition from the standard

[5] I abstract from risk aversion here. Modeling log earnings as concave in the return to education b_i does not change the nature of updating perception.

model remains. The optimal choice of effort solves the first-order condition

$$E_i[b_i] = c_i'(e_i) \tag{1.2}$$

Given that marginal cost is increasing in e_i, the optimal choice of effort will increase if the individual expects higher returns.

The individual's true return to education is determined by the actual average return in the population and by some heterogeneity factors (observable and unobservable characteristics)

$$b_i = \overline{b} + X_i\gamma + \varepsilon_i \tag{1.3}$$

where $\overline{b}$ denotes the actual average return in the population; for example, when the whole economy is doing better, everyone has higher returns. X_i is some observable characteristic that affects one's own return, such as parental wealth (rich or poor background). The last term ε_i captures any heterogeneity unobserved to others and only known by the individual, such as ability. Let ε_i be normally distributed $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$.

Uncertainty about one's own return comes from two sources: imperfect knowledge of the average return ($\overline{b}$) and that of the relationship between his characteristics and his own return (γ). First, assume as in the Bayesian approach that the individual's belief can be described as a prior distribution on the actual average return $\overline{b} \sim N(\overline{b_0}, \sigma_0^2)$. His perceived average return is $E_i[\overline{b}] \equiv \overline{b_0} = \overline{b} + \xi_i$. Second, he does not perfectly observe γ either: $E_i[\gamma] = \gamma + \eta_i$. What determines the effort decision is the individual's expected return

$$\begin{aligned} \underbrace{E_i[b_i]}_{Perceived\ Self} &= \underbrace{E_i[\overline{b}]}_{Perceived\ Average} + E_i[X_i\gamma] + \varepsilon_i & (1.4) \\ &= \underbrace{(\overline{b} + X_i\gamma + \varepsilon_i)}_{Actual\ return\ for\ Self} + \underbrace{(\xi_i + X_i\eta_i)}_{Errors} & (1.5) \end{aligned}$$

With this framework in mind, the next subsection models how individuals update their perceived return after the statistics and role model interventions.

1.2.2 Updating Based on Statistics

Suppose the government receives a noisy signal about the average return $b_G = \bar{b} + \xi_G$. The government's noise is normally distributed $\xi_G \sim N(0, \sigma_G^2)$, i.e. the precision of its signal is $1/\sigma_G^2$. The government provides this statistic b_G to all agents. As the individual sees the government statistic about the average return, he will update his estimate of the average return toward that number. Bayesian updating leads to a normal posterior distribution (DeGroot 1970), with the following mean and variance

$$E_i[\bar{b}|b_G] = \frac{\sigma_G^2}{\sigma_0^2 + \sigma_G^2}\overline{b_0} + \frac{\sigma_0^2}{\sigma_0^2 + \sigma_G^2} b_G \tag{1.6}$$

$$Var_i[\bar{b}|b_G] = \frac{\sigma_0^2 \sigma_G^2}{\sigma_0^2 + \sigma_G^2} \tag{1.7}$$

The individual's posterior perceived average return is a weighted average of his prior and the government statistic. The weight depends on the precision of the government's signal: if σ_G^2 is small, the posterior becomes very close to the statistic. To elaborate, suppose the government's signal b_G comes from random sampling of n observations in the population. The variance of its estimate is $\sigma_G^2 = \frac{s^2}{n}$, where s^2 is an estimate of the population variance. That is, the government's signal is less precise when there is high underlying heterogeneity in the population. I will return to this point later when I discuss heterogeneity in more detail.

According to expression 1.4, agents will also update their own expected return toward the statistic. For people who had initially underestimated the average return ($\overline{b_0} < b_G$), their perceived return increases, resulting in higher effort e. However, one's relative position to the average $E_i[X_i\gamma] + \varepsilon_i$ is still unchanged.

1.2.3 Updating Based on Role Models

The next step is to examine how individuals update their belief after seeing a role model. The role model is simply one observation from the distribution and should not affect beliefs unless this person carries a signal about the return to education. I need to model explicitly what an individual thinks the role model is supposed to signal and why the government chooses to send information in this way. A useful framework in which role models might plausibly affect behaviors is a game between the government and an individual. I formalize individuals' beliefs about the programs and their behaviors in a Perfect Bayesian equilibrium. Recall that each agent is uncertain about the relationship between his characteristics and his own return (γ). Here a welfare-maximizing government receives a signal about this relationship. Through its role model programs, it conveys the information to people.

First, let me specify the observable characteristic X_i to be the individual's type. Suppose there is a continuum of agents of measure 1. There are 2 types H and L (born rich and poor). One's type is unknown to other individuals but observable to the government. Perceived return in expression 1.4 can now be written as

$$E_i[b_i] = E_i[\bar{b}] + \ E(\gamma_H) * H + E(\gamma_L) * L + \varepsilon_i \tag{1.8}$$

where H and L are dummies for each corresponding type. Without loss of generality, I look at an L-type individual throughout the rest of the model. He does not observe γ_L perfectly.

There are 3 states of nature with varying individual return. The "low heterogeneity" state occurs with probability $1 - q$, in which case returns to education do not depend on one's type, i.e. $\gamma_H = \gamma_L = 0$. I call this state "low heterogeneity" since heterogeneity in individual returns in the population comes solely from unobservable (σ_ε^2) rather than from both observable and unobservable characteristics. Alternatively, when there is high heterogeneity (with probability q), this heterogeneity favors one type. In the good state ("good" from the L-type's point of view), being of type L affects the return positively $\gamma_L = \gamma_2 > 0$ while being of type H affects the return negatively $\gamma_H = \gamma_1 < 0$. Symmetrically, in the bad state, $\gamma_L = \gamma_1 < 0$ and $\gamma_H = \gamma_2 > 0$. There is uncertainty about the state. The probability of high heterogeneity is q , and the probability of the good state given high heterogeneity μ_0.

Definition 1 *After the payoff is realized, anyone with income above a threshold $\overline{y}$ is considered to be successful ("role model").*

Consider a static game between 2 players: a welfare-maximizing government and a low-type individual. The timing is as follows. Nature first decides on low or high heterogeneity. Under high heterogeneity, nature decides which type has the higher return (good or bad state for the L-type). The government knows exactly which state it is. The government decides whether to send a role model–example of success, and which type to send. The individual sees the government's action and infers information about the state. He then updates beliefs about the return using Bayes' rule, and chooses the optimal level of effort. The extensive form of the game is displayed in Appendix Figure 1.

1.2.3.1 Individual's Problem

Given the prior beliefs, a low-type individual's expected rate of return is

$$E_i[b_i] = E_i[\overline{b}] + q\mu_0\gamma_2 + q(1-\mu_0)\gamma_1 + \varepsilon_i \tag{1.9}$$

His choice of e solves

$$e_0 = \arg\max\{[E_i[\overline{b}] + q\mu_0\gamma_2 + q(1-\mu_0)\gamma_1 + \varepsilon_i]e - c_i(e)\} \tag{1.10}$$

The efficient choices of education for the low-type are as follows.

In the low heterogeneity state:

$$e^*_{lowhet} = \arg\max\{[E_i[\bar{b}] + \varepsilon_i]e - c_i(e)\} \tag{1.11}$$

In the good state:

$$e^*_{good} = \arg\max\{[E_i[\bar{b}] + \gamma_2 + \varepsilon_i]e - c_i(e)\} \tag{1.12}$$

In the bad state:

$$e^*_{bad} = \arg\max\{[E_i[\bar{b}] + \gamma_1 + \varepsilon_i]e - c_i(e)\} \tag{1.13}$$

1.2.3.2 Government

The welfare-maximizing government receives a fully informative signal of what the state is. It wants to choose an action to signal to the low-type agent about the state so that he can make the efficient choice of effort. The government wants the individual's choice to be as close as possible to the efficient choice. Consider the government's set of possible strategies $g \in \{Nobody, L, H\}$, which represents sending nobody, sending a low-type role model, or sending a high-type role model.

Given a signal status $s \in \{Lowhet, Good, Bad\}$, the government's objective function is

$$\min_{g(s)} |e(g) - e^*(s)| \tag{1.14}$$

where $e(g)$ is the response function of the low-type after observing the government's action g. $e^*(s)$ is the efficient effort level under each of the government's information sets

$$\begin{aligned}
e^*(Lowhet) &\equiv e^*_{lowhet} = \arg\max\{[E_i[\bar{b}] + \varepsilon_i]e - c_i(e)\} && (1.15)\\
e^*(Good) &\equiv e^*_{good} = \arg\max\{[E_i[\bar{b}] + \gamma_2 + \varepsilon_i]e - c_i(e)\} && (1.16)\\
e^*(Bad) &\equiv e^*_{bad} = \arg\max\{[E_i[\bar{b}] + \gamma_1 + \varepsilon_i]e - c_i(e)\} && (1.17)
\end{aligned}$$

1.2.3.3 Equilibrium Concept (Benchmark Model)

Proposition 2 *The following strategies and beliefs constitute a (pure-strategy) Perfect Bayesian equilibrium.*

Government's strategy g(s): $g(Lowhet) = Nobody, g(Good) = L, g(Bad) = H$

Low-type individual's strategy e(g): $e(Nobody) = e^*_{lowhet}, e(L) = e^*_{good}, e(H) = e^*_{bad}$

Individual beliefs at his decision node:

$probability[Lowhet|g = Nobody] = 1$

$probability[Good|g = L] = 1$

$probability[Bad|g = H] = 1$

Proof. We need to show that each strategy is the best response given the other party's strategy, and the beliefs are consistent with Bayes' rule.

At the final decision node, the low-type individual updates his belief on the state after observing the government's move. Given the government's strategy, seeing $g = Nobody$ means that the government signals the low heterogeneity state. Given the government's equilibrium strategy, the individual's posterior belief is $probability[Lowhet|g = Nobody] = 1$. The solution to his maximization problem is e^*_{lowhet} (see expression 1.11). Seeing $g = L$ implies that the state is good, so the individual updates

$$\begin{aligned} prob[good|g &= L] && (1.18)\\ &= \frac{p[L|good] * q\mu_0}{p[L|good]q\mu_0 + p[L|bad]q(1-\mu_0) + p[L|lowhet](1-q)} && (1.19)\\ &= \frac{1 * q\mu_0}{1 * q\mu_0 + 0 * q(1-\mu_0) + 0 * (1-q)} && (1.20)\\ &= 1 && (1.21) \end{aligned}$$

This is good news (also better news than seeing nobody) since he now thinks his return is $E_i[\bar{b}] + \gamma_2 + \varepsilon_i$, and the solution to his maximization problem is e^*_{good}. By similar logic, a high-type role model is bad news to the low-type individual, and leads to e^*_{bad}.

Given the individual's strategy, the government solves its minimization problem as in expression 1.14. It would not deviate from the equilibrium strategy since that would cause the individual to deviate from the efficient choice of effort. ■

1.2.4 Statistics and Role Models

This subsection summarizes the predictions of updating based on statistics and on role models from the previous two subsections, taking into consideration the effect of the combined interventions. Now in equilibrium, sending role models implies high underlying heterogeneity in the population. Recall that the government's statistic b_G is less precise under high heterogeneity. As a consequence, the individual puts less weight on the government's statistic when the government also sends a role model.

The equilibrium strategies in the previous benchmark model can be extended to the following.

Government's strategy:

1. No signal about the average return, low heterogeneity state: do nothing
2. Noisy signal about the average return, low heterogeneity state: provide statistics alone
3. No signal about the average return, high heterogeneity state: send low-type role model if good state, send high-type role model if bad state
4. Noisy signal about the average return, high heterogeneity state: provide statistics, and a role model as in strategy 3.

Low-type individual's strategy:

1. Government does nothing: infer that it is the low heterogeneity state and choose effort accordingly

2. Statistics alone: update perceived returns $E_i[\overline{b}]$ and $E_i[b_i]$, infer that it is the low heterogeneity state, and change effort

3. Low-type role model: infer that it is the good state and increase effort

4. High-type role model: infer that it is the bad state and decrease effort

5. Statistics and low-type role model: update perceived returns, infer that it is the good state but put less weight on the statistics, and change effort

6. Statistics and high-type role model: update perceived returns, infer that it is the bad state and put less weight on the statistics, and change effort

These strategies give us a set of predictions for individual behaviors under the government's statistics and role model programs explored empirically in this paper.

1.2.5 Discussion

There are three important numbers here that will be relevant throughout the rest of the paper. The government has a signal b_G about the actual average return $\overline{b}$, and this is the statistic it provides. People initially do not know $\overline{b}$ and have a perceived return for the average $E_i[\overline{b}] \equiv \overline{b_0} = \overline{b} + \xi_i$. Each individual also has expected return for his own type, i.e. a perceived return to education for himself as in expression 1.4.

When mapped to the empirical setting, this model carries three main insights. First, it predicts that everyone updates toward the statistic about the average return. In particular, this statistic has the same effect for two individuals of different types who start with the same prior about the average return. Second, seeing a role model of the same type (initial background) is good news since this role model reveals to the agent that the heterogeneity is favorable to him. Meanwhile, seeing a role model from a different type is bad news. Third, combining statistics with a role model may signal high heterogeneity in the returns and undermine the impact of statistics.

The sections that follow will explore how people behave empirically when they receive the statistics and role model programs from the government.

1.3 Study Design

Imperfect knowledge of the return, as described in section 1.2's framework, is arguably common in many areas in developing countries. Most of rural Madagascar lies in secluded areas with limited access to outside information, and information on earnings is not covered by the media (radio). According to a pilot study by the Ministry of Education and UNICEF in November 2006, many local parents have difficulties estimating the income associated with various educational attainments. From my survey data, 73% of the respondents report that it is difficult to learn about their peers and neighbors' income; 53% say there are frequent incidences of educated people out-migrating from the village. These observations suggest that rural households might have considerable uncertainty about the returns to education.

1.3.1 Description of the Intervention

To learn about behavioral responses to role models and to returns to education statistics, I designed a field experiment to be carried out by the Ministry of Education (MENRS), with support from the French Development Agency (AFD), UNICEF, and the World Bank. This experiment was implemented as a government program, and households presumably responded in this context. The government programs evaluated here were launched at mid-school year, in February 2007 (the academic calendar in Madagascar runs from September to June).

One important feature of the design was that all participating schools organized a parent-teacher meeting. In the comparison group, parents and teachers discussed during this meeting any typical topics of the school.[6] In the other program schools, in addition to these discussions,

[6]This placebo treatment allows me to avoid confounding the effects of meetings mandated by the government with the effects of statistics or role models. For example, I control for the potential impact coming from teachers

Grade 4 students and their parents received either the "statistics" intervention, the "role model" intervention, or both during this meeting.

First, the "statistics" intervention sought to inform parents of the average returns to education, calculated from the nationwide population. In this treatment, school teachers first presented a few simple statistics based on the 2005 Madagascar Household Survey (Enquête auprès des Ménages) to all Grade 4 students and their parents at a school meeting. For each education level, the audience learned about the distribution of jobs by education, and the mean earnings of 25 year-old Malagasy females and males by levels of education. Then the teacher explained the magnitude of increased income associated with higher educational levels, therefore implying percentage gains or returns to education. Discussion on these statistics lasted about 20 minutes. Parents also received a half-page information card featuring mean earnings by gender and by education, and a visual demonstration of the percentage gain (see sample card in the Appendix Figure 2). I refer to these statistics as the estimated average returns to education (b_G in the model). The numbers reported to families are predicted values from a Mincerian regression of log wage on education levels, age, age squared, and gender. These estimates are robust to adding region of birth fixed effects, but they may be biased due to potential endogeneity problems. Within the experiment, people should still respond to the new information provided that these estimates differ from their prior perception.

Second, the role model program mobilized three types of local role models to share their life story at treated schools. In this context, a "role model" is an educated individual with high income, who grew up in the local school district.[7] Motivated by the theoretical proposition that role models from different backgrounds carry different information, I randomly assigned a role model of poor or rich background to different schools. In practice, it was important to also have a role model of moderate income, possibly the relevant level of success for certain areas. The exact types are

suddenly becoming concerned that the government is paying attention to them, or any motivation to parents from being involved in the school.

[7] Existing definitions of role models vary. A broader definition of a role model refers to any example of success that can inspire others to emulate him (Bala and Sorger 1998, Anderson and Ramsey 1990, Haveman and Wolfe 1995).

defined as follows:

- Role Model type LM (low to medium): a role model of *low-income background* similar to many students in the audience (the father and mother were farmers, the role model went to the village school, for example) who is now *moderately successful* (generating high productivity on farm, or owning a profitable shop)
- Role Model type LH (low to high): a role model of *low-income background* similar to many students in the audience, who is now *very successful* (government official or manager of a big enterprise, earning high income).
- Role Model type HH (high to high): a role model of *high-income background* (father works for government, mother is a school teacher, the role model went to a private preparatory school, for example) who is now *very successful.*

Through a local committee consisting of the school district head, a local NGO leader, and community leaders, MENRS identified and recruited 72 role models of different types between December 2006 and January 2007. This recruitment made no explicit distinction in terms of gender (and it turned out there were 15 female role models). UNICEF Communication Sector provided two days of training on communication skills to the chosen role models so that they could reflect on their personal experiences, write an abbreviated script and practice an oral presentation. This training only guided the stories to cover at least three bases of information: background, educational experiences, and current job and standard of living. These role models then visited selected schools (one visit per school) and shared their life stories with all 4th-graders and their parents at a school meeting. After the role model's speech (about 20 minutes), the meeting proceeded to questions and answers, and open discussions. According to field reports, the topics of discussions were mostly to clarify the role model's story, but the parents also expressed their doubts. They sometimes cited examples of teenagers in their village with a high school degree but without a permanent job.

Most frequently, they justified the difficulties of sending children to schools during periods of food shortages.

Finally, in the combined treatment schools, both of the above interventions took place at the same school meeting. The school teacher first explained returns to education statistics, followed by the role model's exposition.[8] The school meetings were well-attended throughout my study sample.

1.3.2 Sample and Evaluation Design

The study sample consists of 640 public primary schools in 16 districts throughout rural Madagascar (see the map in Figure 1). To arrive at this study sample, I had excluded from the Ministry's roster a few schools that are too far away and extremely difficult to access. These 640 schools were divided into 8 groups of 80 each, to receive different combinations of the aforementioned "statistics" intervention, role model interventions and their combinations. Treatment assignment into the 8 groups was random, stratified by students' baseline test score and AGEMAD treatment status. AGEMAD is an on-going experiment aimed at improving management in the educational system, which also covered all the schools in my sample (Nguyen and Lassibille 2007).[9]

Table 1 describes the precise treatment design and distribution of program components. There are 8 treatment groups (TG0 to TG7). The columns represent the "role model" interventions by type (TG2, TG3, and TG4 received a role model only). The second row represents the "statistics" intervention (TG1 received statistics only). I define "Any Statistics" as the second-row groups of 320 schools, and "Any RM" as the 480 schools in the last three columns. TG5, TG6, and TG7 refer to the combined treatment schools, who received both the statistics and role model treatments. In

[8] Discussion on the statistics took place first to help mitigate the possibility that it might be overshadowed by the role models.

[9] The AGEMAD packaged intervention provides (i) operational tools and training to administrators and teachers, (ii) report cards and accountability meetings with a purpose to improve the alignment of incentives. This cross-cutting design does not a priori restrict external validity since I do not expect any interactions between AGEMAD and the treatments explored in this paper. Indeed, when I test for this interaction, the treatment effects of statistics and role models on test scores do not vary significantly with AGEMAD status.

all program schools, Grade 4 (aged 9-15) is the only treatment cohort. Other cohorts may have been indirectly affected since, for example, parents often have more than one child in the same primary school.

As mentioned earlier, the comparison group (TG0) still received the placebo meeting, but neither the statistics nor the role model interventions. Aside from the design table, 69 other schools were randomly chosen to be "pure control". Those schools did not come into contact with any announcement of the program, nor any baseline survey administered, nor any school meeting organized by the program. I will compare test scores in those schools (the only data available for this group) to those of the regular comparison group to detect any potential effect of simply participating in the experiment: answering survey questions and attending a meeting.

This evaluation design allows me to address the main questions in this paper concerning how individuals respond to information about the returns to education. First, TG1 versus TG0 tells us to what extent providing pure statistical information changes one's perceived returns and schooling decisions. Second, I can measure whether role models are overall effective in changing behaviors, or only someone of the same type has a positive impact on effort. Third, the combined treatment schools help us understand what happens to the impact of statistics when extra information such as a role model is also presented.

1.4 Data and Experimental Validity

1.4.1 Data

1.4.1.1 Background Data

Background information on the schools is available from administrative data collected by the Ministry of Education. In addition, I collected some information on household characteristics in a parent survey in all the schools of my sample. Table 2 presents descriptive statistics at baseline for the sample of schools and households in this study. The average primary school has around 215 students in total enrollment for Grades 1-5, and 30 students in Grade 4. From the baseline test scores, the average student can reach 60% of the competency level they are expected to master. Their families are mostly poor, but I will refer to the top half of income as the relative "rich" in this sample. Less than half of the parents finished primary school, despite their high level of self-reported literacy.

1.4.1.2 Perception of Returns to Education

Evaluation of the statistics and role model interventions rests on three main data sources collected during the experiment: parent surveys, school attendance data, and Grade 4 students' test scores. The first one is subjective data on beliefs, while the latter two give objective measures of schooling. All data, except attendance, were collected for my entire sample at baseline (mid-school year) and ex-post (at the end of the school year). The actual sample turned out to be slightly fewer than 640 schools since some did not have Grade 4 during the study period.

I designed a parent survey to measure the first outcome of interest–perceived returns to education. This data allows us to gain a better understanding of the potential information gap at the household. At the beginning and the end of this project, surveyors visited the homes of Grade 4 students and interviewed either parent of the child (55% of the respondents were mothers). My survey approach follows Jensen (2007) and previous literature to elicit each individual's perceptions of the returns to education, both for the average and for oneself. We want to measure these two numbers since dispersion in perceived returns for oneself reflects both heterogeneity in the actual returns and possible misperception. Dispersion in perceived average returns, on the other hand, would reflect misperception (unless parents do not fully understand the survey question).

The exact survey question on perceived average earnings is as follows:

> "Please estimate the average monthly earnings of current 25 year-old Malagasies

without a primary school degree"

The same question repeats for other scenarios of educational attainment. There are four scenarios in total: no primary education, only primary education, only lower-secondary education, and only high school education. Respondents also gave an estimate of their own child's (the Grade 4 student) income in hypothetical cases of various educational attainments, in response to two consecutive questions:

> "Suppose, hypothetically, that your child were to leave primary school without obtaining the CEPE [primary school degree] and not complete any more schooling. What types of work do you think he/she might (be offered and) choose to engage in when he/she is 25 years old?"
>
> "How much do you think he/she will earn in a typical month at the age of 25?"

The answer to the first question is perceived job type by education. The answer to the second question is perceived earnings by education. From respondents' perceived earnings by education, I calculate the corresponding proportional gain in earnings due to education. I call this proportional gain "perceived returns to education." For each child, there are three levels of perceived returns: additional gain from primary education, lower-secondary education, and high school. For example, the perceived return to lower-secondary school is defined as

$$\frac{Perceived\ Earnings(Secondary) - Perceived\ Earnings(Primary)}{Perceived\ Earnings(Primary)}$$

In this paper, I refer to respondents' estimates of the average returns as "perceived returns for average," and those of the child as "perceived returns for self." I measure for each individual at baseline his perceived returns for self and for average. Endline perceived returns to education is individual-level data, and is name-matched to the baseline perceived returns at 75% match rate.

1.4.1.3 Schooling Outcomes

Attendance rate is measured by the ratio of students present to total enrollment, at the school-grade level. This attendance data was collected by surveyors during unannounced school visits (one visit per school). It is available for only a random subset of schools in the study sample. This variable indicates whether the role model and statistics interventions entice students to exert more effort to attend school.

Lastly, I examine students' achievement through individual test scores. The baseline and post-tests measure children's competency in three materials: mathematics, French, and Malagasy. The baseline test took place in February 2006, and the post-test was administered to the same children (in Grade 4) in June 2007, as part of the AGEMAD project. Due to administrative constraints, only a random sub-sample of 25 students maximum per school took the test. These tests were developed from existing PASEC exams.[10] They cover basic calculations and grammar questions in French and Malagasy, at the level that the students are supposed to master at this stage. These tests are achievement rather than ability tests, so performance can be improved by increasing effort. Test scores are calculated as the percentage of correct answers. Throughout the paper, I report test scores normalized by the control group mean and standard deviation for easier comparisons across different scales. Test score data is name-matched to the baseline perceived returns at 50% match rate.

1.4.2 Experimental Validity

First, as expected given the randomized design, Table 3 shows that baseline test scores, school size, and repetition rate are statistically indistinguishable across the treatment groups and the comparison group. Columns 4 to 6 reassure that pre-existing differences in household data are mostly insignificant as well, with low point estimates. Only baseline perceived returns for average

[10] PASEC (Programme d'Analyse des Systèmes Educatifs de la CONFEMEN) is a program in 15 francophone countries that studies elements of learning for students.

seem to be different in the statistics and role model LM and LH groups.

Second, to minimize the potential bias caused by differential attrition, I tried to measure outcome variables for all the original participants of the program. Both the baseline and endline surveys were administered at the homes of the students just a few months apart. Any attrition is likely to be due to practical constraints of conducting the survey, such as unfavorable weather, rather than endogenous reasons related to the treatments themselves.

Column 1 of Table 4 presents mean attrition rates from the parent survey in all the treatment groups. While the statistics treatment schools have quite high attrition, the differences in mean attrition are not statistically significant. Columns 2 to 6 examine whether the differences in baseline characteristics of schools with high and low attrition vary across treatment groups. These columns report coefficient estimates from regressing each baseline characteristic on attrition rate interacted with treatment group dummies. Only parental education appears different for schools of different attrition rates across some treatment groups.

The post-test was administered to as many of the baseline children as possible. The school director had asked children who no longer attended school to come at the day of the test; and test administrators tried to find the absent students at home. As shown in column 7 of Table 4, attrition rate in test score data is around 0.12 and similar across treatment groups. Column 8 shows that pretest score differences between attriters and stayers are also similar across treatment groups, implying that attrition is not likely to be selected in terms of the pretest.

Due to time and budget constraints, attendance data is available for only a random sub-sample of 176 schools. While the small sample size prevents very precise estimation, there should not be any attrition bias since this sub-sample was randomly chosen.

1.5 Estimation Strategy and Results

The objective of this paper is first to understand better the distribution of initial perceived returns to education, and how perceived returns for self and for average compare to the estimated average return. Then, it examines the effects of statistics and different kinds of role models on perceived returns to education and schooling outcomes. I present the findings below in that order.

1.5.1 Baseline Perceived Returns to Education

I exploit survey data on perceived jobs and perceived earnings as described in section 1.4.1.2. Figure 2A shows the fraction of the respondents who thought their children, with various education levels, would work in a certain sector. If everyone has the correct perception, these fractions should match the empirical job distribution. Most parents associate higher education with jobs in the public sector. In reality, only 33% of high school graduates work for the government while 40% work in commerce and the private sector, a sharp contrast from the beliefs shown in this figure.

Answers to the survey question on perceived earnings reveal huge variation in parents' estimates of the returns at baseline. First, roughly one third of the respondents do not report perceived earnings, answering "Don't Know" to the survey question (even though almost everyone could predict the job type and report his household income). Panel A of Table 5 presents the fraction of the sample not knowing perceived earnings for self and for average at baseline. More respondents claim to know earnings for higher education, perhaps since they can perceive the standard salaries for public sector jobs but not the variable incomes in agriculture. The poor appear somewhat more likely to answer "Don't Know" perceived earnings (also true for the less educated).

Second, conditional on knowing, perceived returns are dispersed; however, the median perception is close to the estimated average return (Mincer estimates from household survey data). Figure 2B plots the kernel densities of the empirical distribution of perceived returns. The wide dispersion in perceived returns for self may reflect both heterogeneity in returns and imperfect information.

It is important to note that perception about the average returns also varies widely, revealing some extent of misperception. Differences between perceived return for self and that for average, i.e. an individual's relative position to the average, are mostly concentrated around zero.

Interestingly, in all cases, the median perceptions are well aligned with the estimated average returns. Despite large dispersion in perceived returns, the median of each distribution is quite close to the vertical line denoting the estimated average return in all the graphs of Figure 2B. Panels B and C of Table 5 report the median perceived returns and standard deviation. For example, the median person in the full sample thinks his marginal return to lower-secondary education is 0.67, i.e. 67% gain in income compared to completing just primary school. It is interesting to note that the poor have higher median perceived returns, though the standard deviation is also higher. These differences in perceived returns reflect underlying differences in perceived earnings. For all education levels, the poor's perceived earnings are consistently lower than those of the rich (see Panel D). Still, the poor expect to gain more from education. Poor people think they earn very little with lower education levels but can increase earnings substantially with higher education. The relatively rich people in my data think they can earn a fair amount even with little education.

Third, dispersion in perception appears larger than dispersion in the actual earnings recorded in household survey data. Panels D and E of Table 5 display the mean and standard deviation of perceived and observed earnings. The mean perceived earnings are higher, which might be reasonable if parents already take into account growth and inflation in estimating children's earnings in the future. The standard deviation in perception is always larger than that in the survey data, for all levels of education. This evidence again suggests some extent of imperfect information about the returns to education.

1.5.2 Estimation Strategy

I first ask whether the statistics program as a whole has an impact on perception and schooling by pooling the "statistics only" and "statistics with role model" schools to be the "any statistics" treatment. I ask the same question about the role model program as a whole. I also discuss the impact of statistics by itself, role model program by itself, and both interventions together. The main specifications are of the following form:

$$Y_{si} = \alpha + \gamma_0 * AnyStat_s + \delta X_{si} + \varepsilon_{si} \quad (1.22)$$

$$Y_{si} = \alpha + \gamma_1 * AnyRM_s + \delta X_{si} + \varepsilon_{si} \quad (1.23)$$

$$Y_{si} = \alpha + \gamma_2 * AnyStat_s + \gamma_3 * AnyRM_s + \gamma_4 * StatRM_s + \delta X_{si} + \varepsilon_{si} \quad (1.24)$$

where Y_{si} is an outcome variable for individual i in school s. AnyStat is a dummy equal to 1 if school s receives any statistics treatment, and similarly AnyRM for any role model treatment. *StatRM* is an indicator for the schools that received the combined interventions, i.e. the intersection of AnyStat and AnyRM. γ's are the coefficients of interest, to be interpreted as the average treatment effect. For example, γ_2 in equation 1.24 is the difference between average Y in statistics (only) schools and that of the comparison schools. $\gamma_2 + \gamma_3 + \gamma_4$ is the effect of receiving both interventions relative to the comparison schools. Standard errors are clustered at the school level. Observations are weighted by the probability of selection, i.e. sampling weights, so that the coefficients of interest are estimated for the population. X_{si} refers to control variables in some specifications. In most cases, I control for the baseline value of the dependent variable, which is likely to have good explanatory power for the dependent variable and improve precision of the coefficient estimates.

For comparison and to evaluate the impact of role models by type, results from the complete specification of all treatment groups are also presented:

$$Y_{si} = \alpha + \sum_k \gamma^k * TG_s^k + \delta X_{si} + \varepsilon_{si} \quad (1.25)$$

where TG_s^k are indicators for whether school s belongs to treatment group k (k=1 to 7) as defined in the design Table 1. Since treatment assignment was random, the errors ε_{si} are orthogonal to treatment group dummies.

Motivated by the model's predictions, I also investigate the treatment effects by initial perception and by type (rich vs. poor). In particular, any schooling improvement due to the statistics treatment should come from individuals whose initial perceived returns were below the estimated average returns (I call this "underestimate"). I test this prediction by running the following regression:

$$
\begin{aligned}
(Posttest - Pretest)_{si} &= \alpha + \gamma_1 AnyStat_s + \gamma_2 AnyRM_s + \gamma_3 StatRM_s + \delta Pretest_{si} && (1.26)\\
&+\lambda_1 AnyStat_s * 1(Under)_{si} + \lambda_2 AnyRM_s * 1(Under)_{si} && (1.27)\\
&+\lambda_3 StatRM_s * 1(Under)_{si} + \theta * 1(Under) + \varepsilon_{si} && (1.28)
\end{aligned}
$$

where 1(under) is a dummy equal 1 if at baseline, the individual perceived the returns to be lower than the statistics provided. The coefficients of interest are $\lambda' s$ on the interaction terms. A positive λ would imply stronger treatment effects on those who had underestimated the returns. Moreover, we expect role model LH to increase the poor's schooling investment, but not the rich's. I run a regression similar to equation 1.26, with all the seven treatment dummies interacted with an indicator for (relatively) rich households. I will present these results in addition to the average treatment effects.

1.5.3 Impact on Perceived Returns

1.5.3.1 Impact of Statistics on Perceived Returns

I first discuss the impact of the statistics intervention on endline perception, as summarized in Table 6. The fraction of respondents not reporting perceived earnings decreased substantially from the baseline to the endline survey. Only 15% of respondents in the comparison group failed to report their perception (answering "Don't Know"), perhaps since the perception questions were asked for the second time at the endline survey. The first column of Table 6 presents results for the probability of not reporting earnings as the dependent variable in equations 1.22, 1.23, 1.24, and 1.25. Since this "Don't Know" outcome is for each education level, these OLS regressions stack up 4 education levels[11] and include education dummies in the regressions. I cannot reject that the statistics treatment has zero impact on the likelihood of "Don't Know" perceived earnings. These results are robust to controlling for baseline Don't Know status and robust to running probit. Results are similar for Don't Know perceived earnings for average (not shown). While this finding is surprising, it has the advantage that the results below are not biased by selective non-response.

The first-order question is whether people update perceived returns toward the statistics, as predicted by the model. I find that providing statistics reduces the gaps between baseline perceived returns and the estimated average returns that we saw in Figure 2B. Figure 3 and Figure 4 plot the kernel densities of endline perceived returns in statistics schools and comparison schools. Eyeballing these graphs, we see that the statistics schools' distribution tends to be more concentrated around the vertical line of estimated average return. In addition, dispersion in the empirical distribution of perceived returns also decreases. I reject the hypothesis that the variance of perceived returns is the same in the comparison schools and statistics schools (p-value from the variance ratio test is very close to zero). Table 6 confirms this conclusion by presenting estimates of equations 1.22, 1.23, 1.24, and 1.25. The outcome variable in columns 2 and 3 is now the absolute distance between endline perceived returns and the estimated average returns. Since the perceived return outcome is for three levels of education,[12] these OLS regressions include education dummies. Panel A shows that providing only statistics significantly decreases the gap by 0.149 from the control group average of 0.68. This reduction in the gap is similar for perceived returns for self and for average. Note that combining statistics with role models undoes this effect: the combined treatments have no impact on perceived returns. As a result, the average impact of any

[11] No primary education, only primary education, only lower secondary education, and only high school education.

[12] Primary education, lower secondary education, and high school.

statistics on the gap is not significant (Panel B). I will return to this point later when I discuss role models.

Consistent with the theory, providing statistics leads individuals to update their perceived returns for average rather than their relative position to the average. Column 4 of Table 6 shows that the statistics treatment does not affect how far individuals think their returns are from their perceived average. The coefficient estimate on AnyStatistics is close to zero and insignificant. Given that people may have had imperfect information about the average returns, this evidence suggests that they update their perceived returns for average toward the statistics and also update their perceived returns for self by the same amount.

The above estimations on endline perceived returns potentially face a sample selection concern since the "Don't Know" population is equivalent to missing data. However, the rate of "Don't Know" is similar in the statistics schools and the comparison schools. People who did not report perceived earnings have similar baseline characteristics across treatment groups, except for some cases of slight income differences. Thus, selection is not likely to severely bias the impact of statistics. In columns 5 and 6 of Table 6, I also restrict the sample to those who reported their perceptions in both survey rounds and find similar results. The impact of statistics on perception does not seem to be driven by the "Don't Know" population.

1.5.3.2 Impact of Role Models on Perceived Returns

Overall, the role model interventions have small effects on perception. The coefficients on role models in columns 2 and 3 of Table 6 are close to zero and statistically insignificant. On average, role models of any type decrease the gap between perceived return for self and the estimated average return by 0.013, but this estimate is not statistically distinguishable from zero (Panel A). Panel D reports the results for each type of role model. In the entire sample, none of the role model types has a statistically significant impact on the gap in perceived returns.

The only type that may affect perception is role model LH. Recall that type LH is a role model from the poor family background who has become very successful. When I control for baseline perceived returns in the restricted name-matched sample, I find that role model LH combined with statistics decreases the gap between perceived return for average and the estimated average return by 0.13. Column 8 presents the treatment effects for the sample of poor families. Role model LH, alone and when combined with statistics, decreases the gap for the poor. However, this effect is not statistically different from the effect of role model HH.

Consistent with the theory, role models undermine the impact of statistics in general. Throughout different samples of the regression in Panel A, the coefficient estimates on the combined treatment are of similar magnitude but of the opposite sign as the coefficient on statistics. In column 2, for example, statistics alone decreases the gap by 0.149 while statistics and role model changes the gap by -0.149 - 0.013 + 0.159 = -0.003, i.e. reverting the impact of statistics toward zero. This finding suggests that households give less weight to the statistics in their updating when it is accompanied by a role model since the role model signals high heterogeneity in the returns and, therefore, low precision of the statistics.

1.5.4 Impact on Education Investment

1.5.4.1 Impact of Statistics on Education Investment

Given that at least some people update their perceived returns toward the statistics, we would expect test scores and school attendance to also respond. In particular, providing statistics should increase test scores for children with low initial perceived returns.

The statistics program's impact on test scores is consistent with these theoretical predictions. Recall that these tests are achievement tests, and performance can be improved by exerting effort. Since test scores are persistent, I control for each child's pretest score and report results from equations 1.22, 1.23, 1.24, and 1.25 in column 2 of Table 7. The dependent variable is improvement in normalized test scores. As shown in Panel A, the statistics (only) program increases test score improvement by 0.20 standard deviations, conditional on what can be explained by the pretest. This coefficient estimate is both statistically and economically significant. When we pool all statistics schools together, test scores are on average 0.10 standard deviations higher than those in schools without the statistics intervention. As a robustness check, the simple difference regression that does not control for pretest score gives very similar coefficients, but less precisely estimated.

Individuals in our sample started with dispersed perceived returns, so we would expect the treatment effects on test scores to be heterogenous. The theory predicts that schooling improvement should come from individuals whose initial perceived returns were below the actual average returns. For the subsample of test score data that can be name-matched to the baseline perceived return data, column 6 presents the result from estimating equation 1.26. Consistent with the theoretical prediction, the coefficient estimate of the statistics interaction term is positive and statistically significant. In particular, providing statistics increased test scores for those who initially underestimated the returns by 0.365 standard deviations (column 4), while decreasing test scores for those who overestimated the returns by 0.223 standard deviations (column 5). Since the error terms are orthogonal to the treatment indicators, we can read these coefficients as the treatment effects for those initially overestimating or underestimating the returns. A concern in interpreting these results as Bayesian updating is that "underestimating" might be correlated with other factors affecting test scores. For example, everyone responded by increasing effort, but those who overestimated the returns are somehow also those with negative productivity of effort.

The results on school attendance are also indicative of the impact of statistics. Column 1 of Table 7 presents the results for school-level student attendance as the dependent variable in the main estimating equations. Since these attendance rates are class averages rather than individual-level data, observations are also weighted by class size. The regression outputs are for a random sub-sample of 176 schools for which attendance data post-treatment is available. I find that providing solely statistics increases average attendance by 7.846 percentage points, though imprecisely estimated due to the small sample size. Pooling all the statistics treatment together improves power. As reported in Panel B under column 1, schools that received any statistics treatment have on average 3.5 percentage points higher attendance rate in Grade 4, compared to the control group mean of 85.6%. The results are robust to controlling for baseline school characteristics and student attendance one school year before the treatment.

1.5.4.2 Impact of Role Models on Education Investment

Since role models have small effects on perception, their overall impact on schooling investment is also small, as expected. The average test score is practically the same in the schools that received any role model and those without role models (Panel C of Table 7). However, the role model's background seems to matter, as discussed in section 1.5.3.2. As shown in Panel D, column 2 of Table 7, role model type LH increased test score significantly by 0.17 standard deviations. Meanwhile, the estimated impact of role model HH is close to zero. I reject the null hypothesis that role model type LH and type HH have equal coefficients (F-test p-value of 0.06).

The theory predicts that role model LH should have a positive influence on test scores for students of poor family background. Column 9 reports estimation results from running a regression similar to equation 1.26, with all the seven treatment dummies interacted with an indicator for rich households. Role model LH increases average test score by 0.27 standard deviations for the poor, but has practically little impact (statistically insignificant) on children from richer families. In addition, I test for a difference-in-difference effect of role model LH vs. HH for the rich vs. poor. However, with the large standard errors around the coefficient estimate for RM HH, I cannot reject the difference between this coefficient and the coefficient on RM LH (the p-value is 0.34).

1.5.5 External Validity

Since all the comparison schools in my experiment organized a "placebo" school meeting, the treatment effects I have measured may be safely attributed to the statistics or role model treatments themselves. In addition, I have another group of "pure control" schools, which never came into contact with the program nor with the perceived returns survey. The average test score in those schools is very similar to that of the regular comparison schools. This evidence suggests that simply participating in the experiment, i.e. organizing a meeting and being surveyed, does not seem to have a significant impact on behaviors.

1.6 Conclusion

Information about the returns to education is essential to school decision making. In this paper, I document that households in rural Madagascar have imperfect information about the returns to education. I set up a field experiment to evaluate two approaches to alleviating imperfect information: providing statistics on the returns to education, and sending a role model to the school. Consistent with a model of belief formation, the results suggest that households update their perceived returns after receiving the statistics and change schooling decisions accordingly. The statistics intervention improved test scores by 0.37 standard deviations for those who had initially "underestimated" the returns, and reduced test scores for those who had "overestimated" the returns. As the theory predicts (and different from much of the policy discussion on role models), role models have an ambiguous effect. They signal heterogeneity in the return, reducing the impact of statistics toward zero. However, they increase schooling effort of those children whose family background matches theirs.

Overall, the statistics program is a very cost-effective way to entice children to attend school and improve test scores. Even though providing statistics decreased perceived returns to education for some children and increased perceived returns for others, it led to an increase in schooling on average. This program cost 0.08 USD per student but increased student attendance by 3.5 percentage points and improved test scores by 0.20 standard deviations after three months. This implies a program cost of 2.30 USD for an additional year of schooling and 0.04 USD for additional 0.10 standard deviations in test scores, more cost-effective than previous interventions evaluated in a randomized experiment (Jameel Poverty Action Lab 2005). Providing deworming drugs, to date the most cost-effective way to increase attendance, costs 3.50 USD for an additional year of schooling (Miguel and Kremer 2004). In terms of improving test scores, the extra-teacher balsakhi program in India costs 0.67 USD per improvement of 0.10 standard deviations (Banerjee, Duflo, Cole, and Linden 2007).

As a broader implication, my paper suggests that households respond to changes in perceived returns when making schooling decisions. Thus, they would probably respond if the actual market returns improved. As suggested in Foster and Rosenzweig (1996), market interventions to raise the market returns to education may be very effective in increasing education.

There are several caveats and as yet unexplored issues that I plan to address in continuing research. One main caveat is that the results presented here reveal only the short-run and direct impacts. It will be useful to follow the students in my sample to see if their enrollment in later school years increases because of the treatment, and if they will complete the primary cycle and go on to secondary school after Grade 5. Collecting outcome data for other children in the same school would also shed light on spillovers of information. Given the strong direct effects of statistics, it is important to investigate in future research whether and how strongly such information spreads to affect schooling decisions of other children in the same family, and of those in other families through teachers or parents' social networks.

References

ANDERSON, R., AND P. RAMSEY (1990): *Women in Higher Education: Changes and Challenges*chap. Women in Higher Education: Development through Administrative Mentoring, pp. 283–285. New York: Praeger.

AVERY, AND KANE (2004): *College Choices: The Economics of Where to Go, When to Go, and How to Pay for It*chap. Student Perceptions of College Opportunities: The Boston COACH Program. University of Chicago Press.

BALA, V., AND G. SORGER (1998): "The Evolution of Human Capital in an Interacting Agent Economy," *Journal of Economic Behavior & Organization*, 36, 85–108.

BANERJEE, A., E. DUFLO, S. COLE, AND L. LINDEN (2007): "Remedying Education: Evidence from Two Randomized Experiments in India," *Quarterly Journal of Economics*, 122(3).

BETTS, J. (1996): "What Do Students Know about Wages? Evidence from a Survey of Undergraduates," *Journal of Human Resources*, 31(1), 27–56.

BILS, M., AND P. KLENOW (2000): "Does Schooling Cause Growth?," *American Economic Review*, 90(5), 1160–1183.

CARD, D. (1999): "The Causal Effect of Education on Earnings," in *Handbook of Labor Economics, Vol. 3A*, ed. by O. Ashenfelter, and D. Card, pp. 1801–63. North Holland.

DEE, T. (2004): "Teachers, Race and Student Achievement in a Randomized Experiment," *The Review of Economics and Statistics*, 86(1), 195–210.

DOMINITZ, AND MANSKI (1996): "Eliciting Student Expectations of the Returns to Schooling," *Journal of Human Resources*, 31(1), 1–26.

DUPAS, P. (2006): "Relative Risks and the Market for Sex: Teenagers, Sugar Daddies, and HIV in Kenya," mimeo, Dartmouth College.

FOSTER, A. D., AND M. R. ROSENZWEIG (1996): "Technical Change and Human-Capital Returns and Investments: Evidence from the Green Revolution," *American Economic Review*, 86(4), 931–953.

GLEWWE, P., AND M. KREMER (2005): "Schools, Teachers, and Education Outcomes in Developing Countries," Handbook on the Economics of Education (forthcoming).

HAVEMAN, AND WOLFE (1995): "The Determinants of Children's Attainment: A Review of Methods and Findings," *Journal of Economic Literature*, 32, 1829–78.

JAMEEL POVERTY ACTION LAB (2005): "Fighting Poverty: What Works?," Bulletin Issue 1, MIT, Cambridge, MA.

Jensen, R. (2007): "The Perceived Returns to Education and the Demand for Schooling," Mimeo, Brown University.

Manski, C. F. (1992): *Studies of Supply and Demand in Higher Education*chap. Adolescent Econometricians: How Do Youth Infer the Returns to Education? University of Chicago Press.

Miguel, E., and M. Kremer (2004): "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities," *Econometrica*, 72(1), 159–218.

Nguyen, T., and G. Lassibille (2007): "Improving Management in the Educational System in Madagascar: An Experimental Approach," Mimeo, MIT.

Oreopoulos, P. (2003): "Do Dropouts Drop Out too Soon? International Evidence from Changes in School-Leaving Laws," NBER Working Paper 10155.

Psacharopoulos, G. (1985): "Returns to Education: A Further International Update and Implications," *Journal of Human Resources*, 20(4), 583–604.

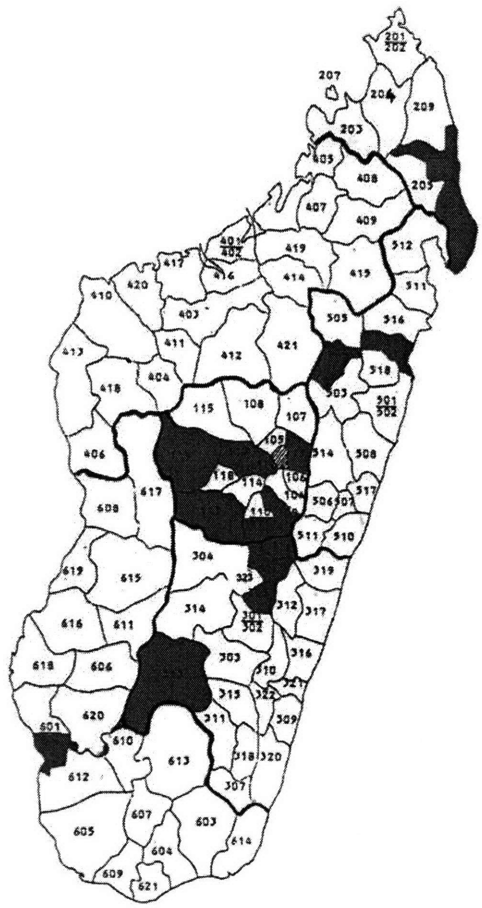
——— (1994): "Returns to Investments in Education: a Global Update," *World Development*, 22(9), 1325–1343.

Ray, D. (2004): "Aspirations, Poverty and Economic Change," Bureau for Research in Economic Analysis of Development (BREAD) Policy Papers No. 442.

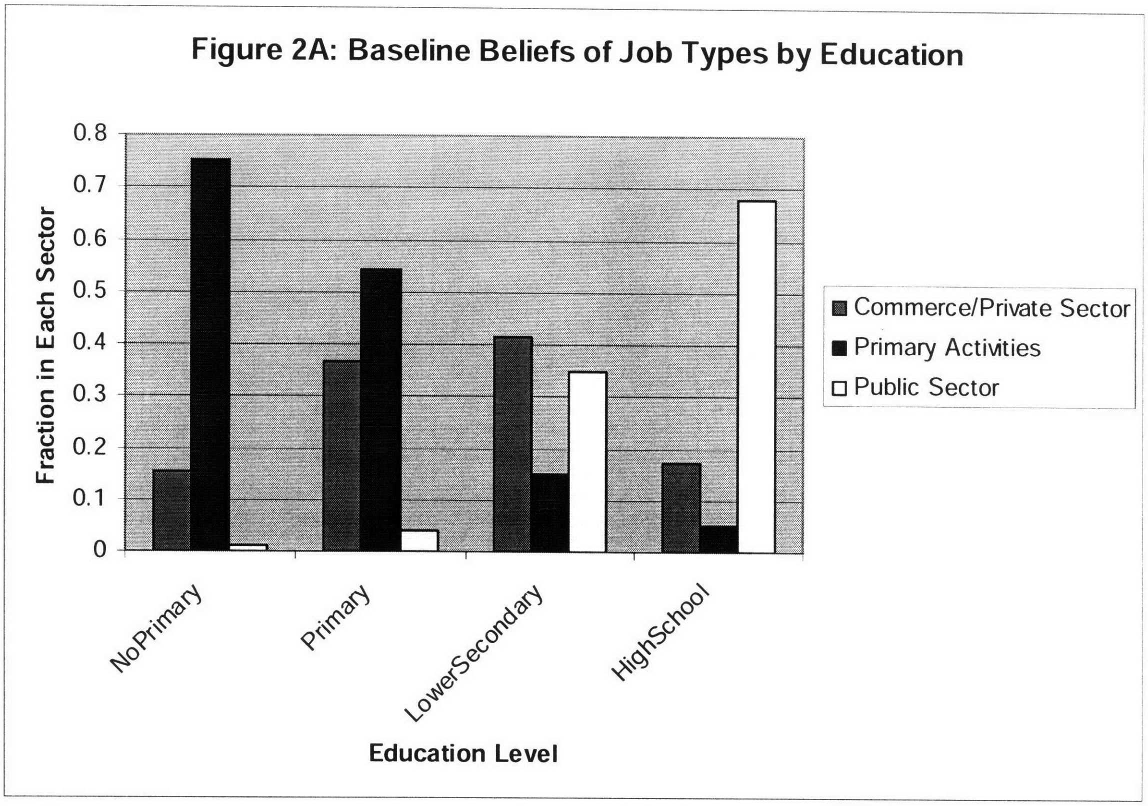
Tan, J.-P. (2005): "Managing for Results in Primary Education in Madagascar: Evaluating the Impact of Selected Interventions," Concept Note, World Bank.

Wilson, W. J. (1987): *The Truly Disadvantaged.* University of Chicago Press.

Figure 1: School District Map of Madagascar

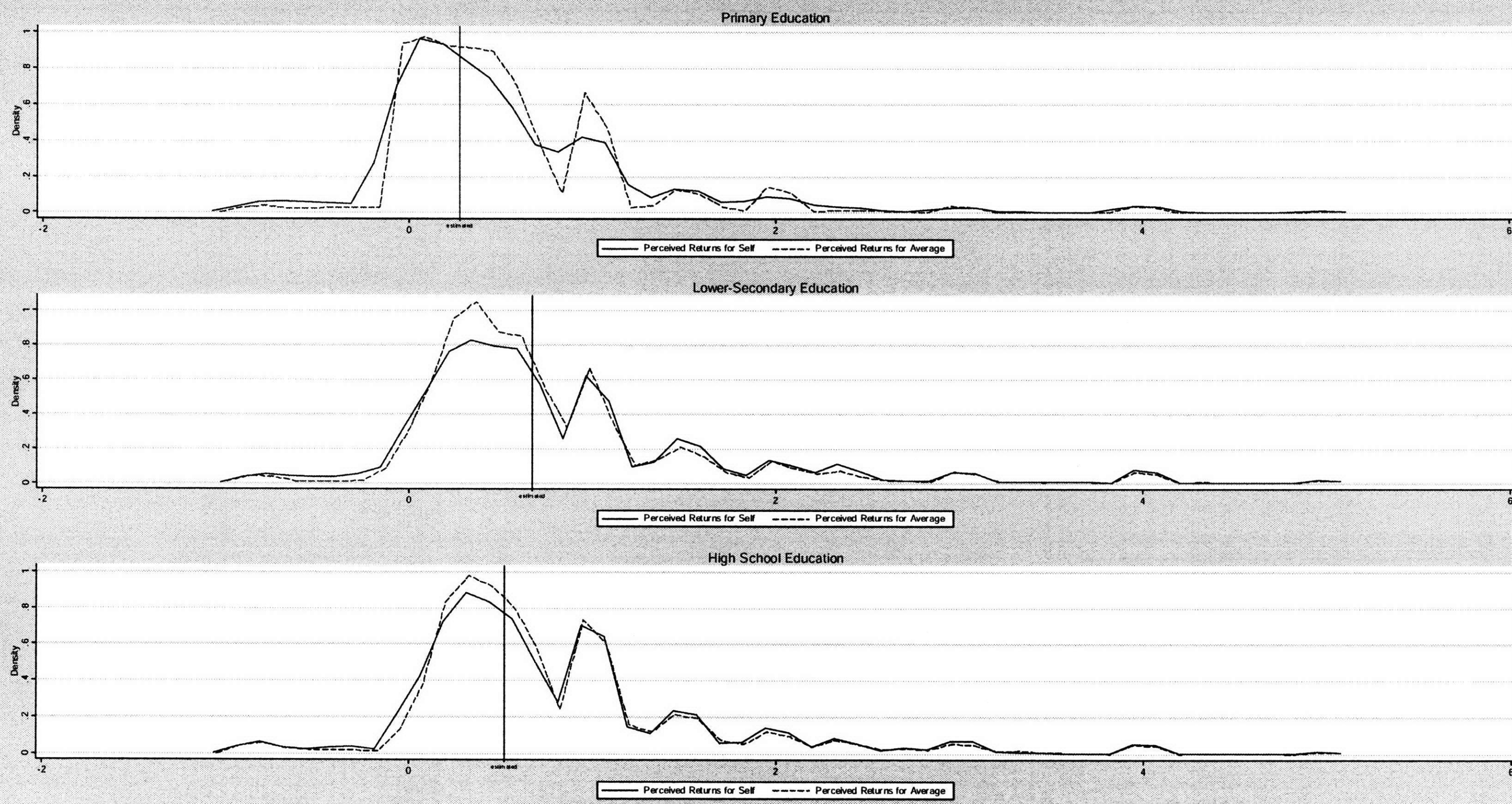


Notes: The shaded areas are the 16 school districts in my study sample



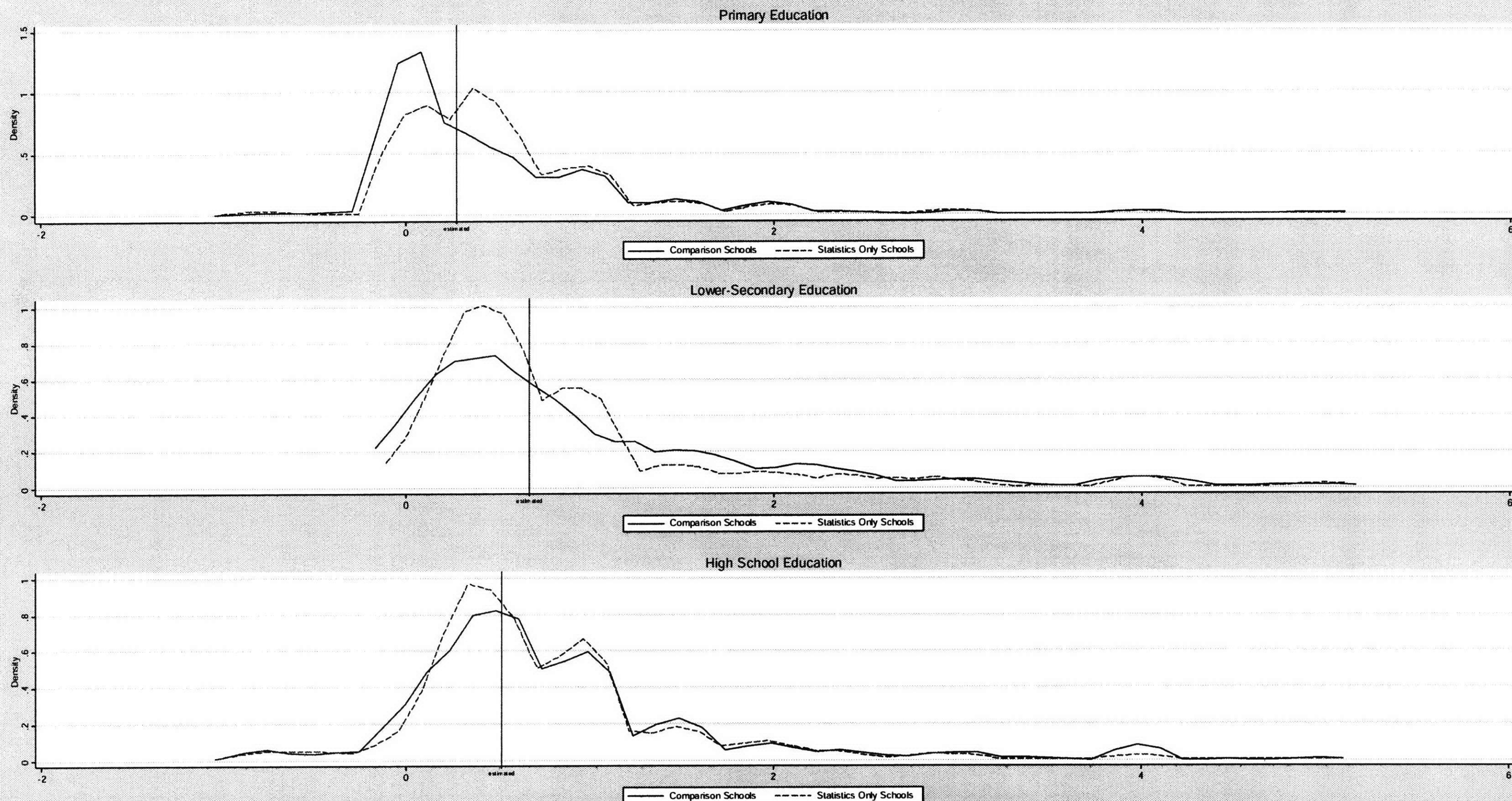
Notes: "Primary Activities" include agriculture, cattle-raising, fishing, and forestry activities. The survey questions on beliefs of job types ask respondents for the sector in which they think their child could work at the age of 25, under different hypothetical scenarios of educational attainment.

Figure 2B: Baseline Density of Perceived Returns for Self and for Average



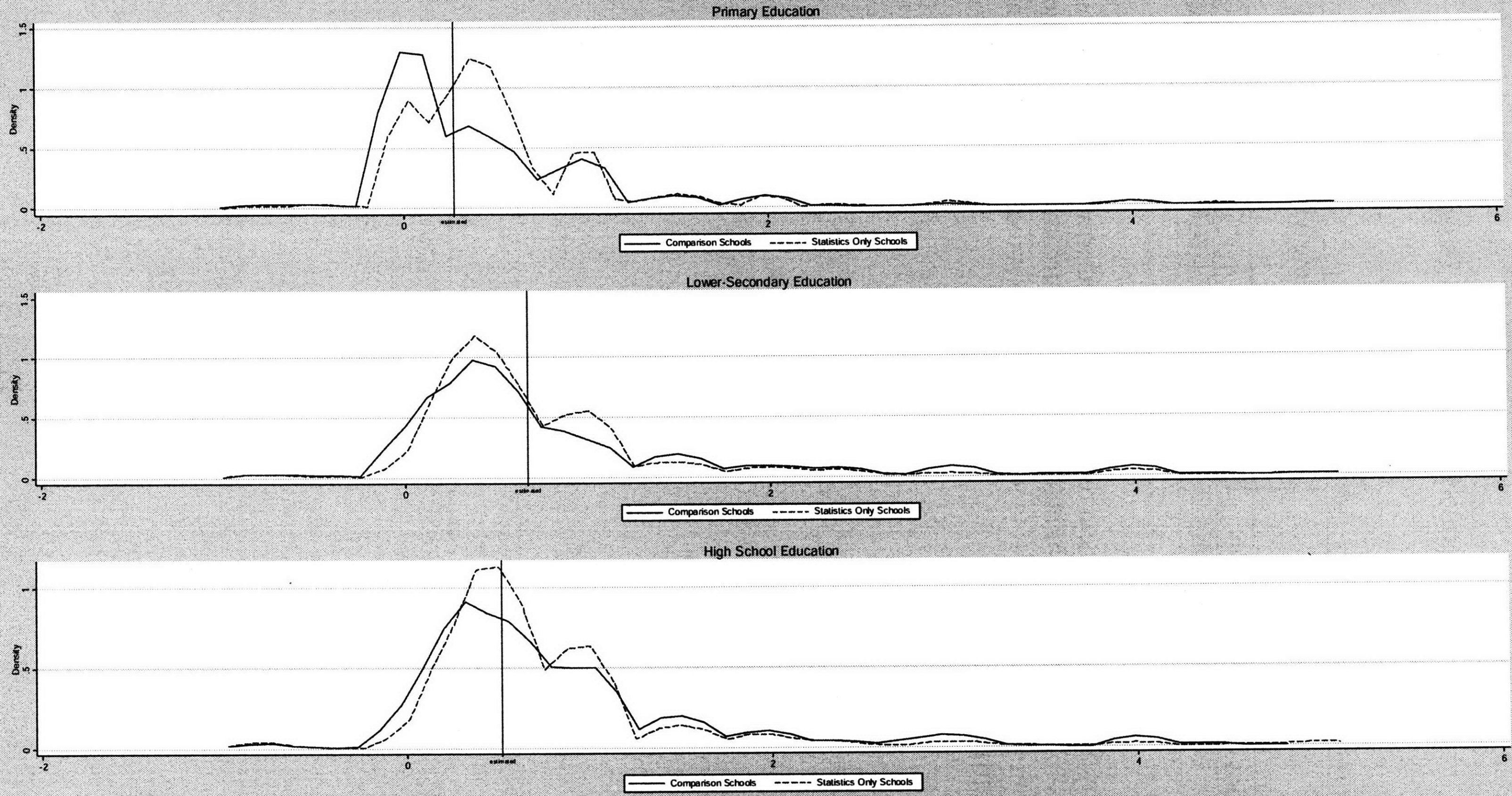
Note: Each panel graphs the kernel density of baseline perceived returns to each level of education.
The solid line represents perceived returns for self, and the dashed line represents those for average.
The vertical lines mark the estimated average returns. Observations are weighted by the sampling probability.
The few values above 5 are not shown.

Figure 3: Endline Density of Perceived Returns for Self Comparison Schools and Statistics Schools



Note: Each panel graphs the kernel density of endline perceived returns to each level of education. The solid line represents the comparison schools, and the dashed line represents statistics schools. The vertical lines mark the estimated average returns. Observations are weighted by the sampling probability. The few values above 5 are not shown.

Figure 4: Endline Density of Perceived Returns for Average Comparison Schools and Statistics Schools



Note: Each panel graphs the kernel density of endline perceived returns to each level of education.
The solid line represents the comparison schools, and the dashed line represents statistics schools.
The vertical lines mark the estimated average returns. Observations are weighted by the sampling probability.
The few values above 5 are not shown.

Table 1: Experimental Design

		Role Model Treatment		
	No RM	**Type LM**	**Type LH**	**Type HH**
No Statistics	TG0: 80 schools	TG2: 80 schools	TG3: 80 schools	TG4: 80 schools
Statistics	TG1: 80 schools	TG5: 80 schools	TG6: 80 schools	TG7: 80 schools

Notes: "Any Statistics" Treatment refers to the second-row groups of schools. "Any RM" refers to the last three columns. The combined treatment refers to groups 5, 6, and 7 where both the statistics and role model interventions took place. TG0 denotes the regular control group.

Table 2: Descriptive Statistics

	Mean	Std. Dev.	Obs
	(1)	(2)	(3)
Panel A: School Variables			
Baseline Test Score	60.67	11.84	640
Enrollment	215.01	153.57	640
Repetition Rate	0.19	0.11	640
Number of Sections in School	5.51	2.03	640
Number of Classrooms in School	4.01	2.04	640
Panel B: Household Variables			
Respondent Age	42.48	9.46	17158
Female Respondent	0.55	0.50	17158
Household Size (in persons)	6.60	2.11	17158
Respondent Literacy: Read and Write	0.88	0.33	17158
Respondent Literacy: Calculations	0.85	0.35	17158
Mother Primary Education	0.41	0.49	17158
Mother Secondary or Higher Education	0.01	0.10	17158
Father Primary Education	0.44	0.50	17158
Father Secondary or Higher Education	0.03	0.16	17158
Household Monthly Income (in Ar)	49703.97	90615.65	16754

Notes: 2000 Ar~ 1 USD

Omitted Education category: No primary school

Table 3: Baseline Differences across Groups

	Dependent variables								
	Baseline Test Score	Enrollment	Repetition Rate	Mother Has Primary Education	Father Has Primary Education	Household Monthly Income (in Ar)	Fraction Not Reporting Perceived Own Earnings	Perceived Returns for Self	Perceived Returns for Average
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
TG1 = Statistics	-0.509	-9.437	-0.017	-0.038	-0.028	9557.45	-0.019	-0.079	-0.204
	(1.76)	(22.28)	(0.02)	(0.04)	(0.04)	(10047.76)	(0.08)	(0.09)	(0.065)**
TG2 = RM LM	0.03	44.375	0.012	-0.068	-0.063	855.201	0.075	-0.106	-0.155
	(1.83)	(24.40)	(0.02)	(0.032)*	(0.03)	(5122.38)	(0.08)	(0.09)	(0.074)*
TG3 = RM LH	0.163	8.663	0.006	0.014	-0.003	823.111	-0.144	-0.12	-0.143
	(1.78)	(20.32)	(0.02)	(0.03)	(0.03)	(4468.32)	(0.058)*	(0.08)	(0.069)*
TG4 = RM HH	0.268	13.613	0.02	-0.003	0.009	3967.85	0.068	0.073	0.029
	(2.01)	(28.50)	(0.02)	(0.03)	(0.03)	(4587.45)	(0.07)	(0.12)	(0.13)
TG5 = Statistics+RM LM	0.076	7.738	-0.019	-0.062	-0.066	-2095.79	-0.01	-0.084	-0.086
	(1.85)	(21.63)	(0.02)	(0.03)	(0.033)*	(4332.30)	(0.07)	(0.09)	(0.08)
TG6 = Statistics+RM LH	-0.254	26.888	0.012	-0.009	-0.017	10435.82	-0.045	-0.013	-0.037
	(1.95)	(21.60)	(0.02)	(0.04)	(0.03)	(8248.42)	(0.07)	(0.09)	(0.07)
TG7 = Statistics+RM HH	0.299	-9.687	0.02	-0.049	-0.051	295.55	-0.067	-0.1	-0.076
	(2.11)	(18.77)	(0.02)	(0.04)	(0.03)	(4736.78)	(0.07)	(0.09)	(0.08)
Constant	60.66	204.74	0.19	0.43	0.47	46591.49	0.41	0.78	1.11
	(1.365)**	(14.429)**	(0.014)**	(0.024)**	(0.023)**	(3,592.01)**	(0.054)**	(0.068)**	(0.061)**
Observations	640	640	640	17158	17158	16754	68632	31293	31306
F-stat (joint significance)	0.05	1.19	1.53	1.85	1.67	0.76	3.52	0.87	2.31

Notes: This table presents OLS results from regressing baseline school characteristics on different treatment group dummies.
Omitted category: control group.
Columns 1-3: school-level data, and robust standard errors in parentheses
Columns 4-6: individual-level data; Columns 7-9: individual data by education; standard errors in parentheses are clustered at the school level
Exchange rate: 2000 Ar~ 1 USD
Omitted education category: No primary school
* significant at 5%; ** significant at 1%

Table 4: Attrition in Endline Data

	Endline Perceived Returns						Post-test	
		Baseline Characteristics across Schools of Differential Attrition						*Diff (Attriters – Stayers)*
	Percent Attrition	Mother Has Primary Education	Father Has Primary Education	Household Monthly Income (in 10000 Ar)	Perceived Returns for Self	Perceived Returns for Average	Percent Attrition	Pretest Score
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
TG1 = Statistics	0.30	-0.097	-0.103	5.722	-0.338	-0.555	0.13	-0.182
	(0.11)	(0.15)	(0.13)	(4.35)	(0.42)	(0.277)*	(0.02)	(0.15)
TG2 = RM LM	0.19	-0.161	-0.182	3.364	-0.423	-0.387	0.14	0.037
	(0.07)	(0.12)	(0.110)+	(2.82)	(0.42)	(0.31)	(0.02)	(0.13)
TG3 = RM LH	0.15	-0.221	-0.179	1.033	-0.439	-0.391	0.11	0.047
	(0.07)	(0.108)*	(0.086)*	(2.32)	(0.35)	(0.28)	(0.01)	(0.16)
TG4 = RM HH	0.10	-0.195	-0.257	-0.2	0.058	-0.475	0.12	0.038
	(0.05)	(0.14)	(0.17)	(3.04)	(0.58)	(0.46)	(0.02)	(0.16)
TG5 = Statistics+RM LM	0.01	-0.205	-0.197	0.315	-0.31	0.125	0.12	-0.057
	(0.05)	(0.105)+	(0.083)*	(2.25)	(0.34)	(0.32)	(0.02)	(0.18)
TG6 = Statistics+RM LH	0.06	-0.014	-0.02	-5.272	-0.414	-0.624	0.10	-0.19
	(0.03)	(0.14)	(0.12)	(4.12)	(0.56)	(0.53)	(0.01)	(0.15)
TG7 = Statistics+RM HH	0.05	-0.442	-0.344	-2.663	0.004	-0.595	0.12	-0.13
	(0.02)	(0.204)*	(0.174)*	(2.66)	(0.46)	(0.42)	(0.01)	(0.17)
Control Group	0.09						0.13	
	(0.08)						(0.02)	
Observations	619	17029	17029	16627	31097	31125		13162

Notes:
Attrition rates in endline perceived returns data are at the school level.
Column 1: school attrition rates are weighted by class size in the baseline data and sampling probability. Robust standard errors in parentheses.
Columns 2-6 report the results from regressing each baseline characteristic on attrition rate interacted with treatment group dummies
Standard errors in parentheses are clustered at the school level. Omitted categories: Control group.
Attrition rates in posttest data are at the individual level. Column 8 reports the pretest score differences between attriters and stayers across groups
+ significant at 10%; * significant at 5%; ** significant at 1%

Table 5: Perceived Earnings and Perceived Returns to Education at Baseline

	Full sample		Rich		Poor	
	Perceived (Self)	Perceived (Average)	Perceived (Self)	Perceived (Average)	Perceived (Self)	Perceived (Average)
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Fraction Not Reporting Perceived Earnings						
No Primary	0.37	0.38	0.35	0.33	0.38	0.41
Primary	0.35	0.37	0.32	0.32	0.35	0.39
Lower Secondary	0.30	0.34	0.27	0.29	0.31	0.37
High School	0.28	0.34	0.24	0.29	0.30	0.37
Panel B: Median Perceived Returns						
Primary	0.33	0.38	0.33	0.33	0.50	0.43
Lower Secondary	0.67	0.60	0.50	0.50	0.67	0.67
High School	0.67	0.67	0.63	0.65	0.67	0.67
Panel C: Standard Deviation of Perceived Returns						
Primary	1.52	1.14	1.40	1.05	1.60	1.21
Lower Secondary	1.85	1.66	1.50	1.38	2.10	1.89
High School	1.80	1.63	1.55	1.52	1.98	1.73
Panel D: Mean and Standard Deviation of Perceived Monthly Earnings (in Ar.)						
No Primary	46182	39117	53923	44325	38301	33509
	(56452)	(42582)	(59614)	(43248)	(51002)	(40611)
Primary	65352	55455	73311	62602	57506	47912
	(77133)	(62336)	(78101)	(65224)	(75178)	(57625)
Lower Secondary	109712	90920	117742	97972	101348	83185
	(120973)	(91504)	(124307)	(92108)	(115293)	(88921)
High School	197281	161548	207859	172628	185960	149695
	(219231)	(151763)	(225197)	(158897)	(210685)	(142717)
Panel E: Mean and Standard Deviation of Observed Monthly Earnings in Household Survey						
No Primary		20699				
		(30649)				
Primary		23812				
		(52465)				
Lower Secondary		54025				
		(77611)				
High School		100333				
		(157325)				

Notes: Earnings measured in Ariary (Ar). Exchange rate: 2000 Ar~ 1 USD

Table 6: Impact on Endline Perceived Returns to Education

	Dependent variables								
		Gap[(#)] (Entire sample)			Gap (Matched sample)				
	Not Report	Self	Average	Self – Average	Self	Average	Rich	Poor	Difference (Rich – Poor)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A									
AnyStatistics	0.054	-0.149	-0.158	-0.024	-0.23	-0.21	-0.123	-0.374	0.261
	(0.04)	(0.081)+	(0.085)+	(0.05)	(0.086)**	(0.086)*	(0.11)	(0.099)**	(0.126)*
AnyRM	0.089	-0.013	-0.062	0.009	-0.011	-0.068	0.069	-0.128	0.202
	(0.032)**	(0.07)	(0.07)	(0.04)	(0.08)	(0.06)	(0.09)	(0.09)	(0.104)+
Statistics+RM	-0.074	0.159	0.202	0.003	0.237	0.228	0.109	0.409	-0.313
	(0.05)	(0.096)+	(0.097)*	(0.06)	(0.106)*	(0.099)*	(0.14)	(0.118)**	(0.147)*
Panel B									
AnyStatistics	-0.003	-0.03	-0.007	-0.022	-0.053	-0.04	-0.051	-0.052	0.002
	(0.03)	(0.05)	(0.04)	(0.03)	(0.06)	(0.05)	(0.07)	(0.06)	(0.07)
Panel C									
AnyRM	0.053	0.061	0.033	0.01	0.106	0.043	0.128	0.065	0.061
	(0.025)*	(0.05)	(0.05)	(0.03)	(0.061)+	(0.06)	(0.075)+	(0.07)	(0.09)
Panel D									
TG1 = Statistics	0.054	-0.149	-0.158	-0.024	-0.23	-0.21	-0.123	-0.374	0.261
	(0.04)	(0.081)+	(0.085)+	(0.05)	(0.086)**	(0.086)*	(0.11)	(0.099)**	(0.126)*
TG2 = RM LM	0.15	0.052	-0.049	0.046	0.067	-0.06	0.188	-0.089	0.284
	(0.047)**	(0.09)	(0.08)	(0.05)	(0.10)	(0.08)	(0.13)	(0.12)	(0.138)*
TG3 = RM LH	0.04	-0.034	-0.054	-0.02	-0.07	-0.079	-0.015	-0.156	0.147
	(0.04)	(0.09)	(0.08)	(0.05)	(0.10)	(0.08)	(0.10)	(0.13)	(0.13)
TG4 = RM HH	0.059	-0.058	-0.081	-0.002	-0.037	-0.066	0.032	-0.144	0.177
	(0.04)	(0.09)	(0.09)	(0.05)	(0.10)	(0.08)	(0.14)	(0.10)	(0.13)
TG5 = Statistics+RM LM	0.041	0.057	0.052	-0.015	0.027	-0.043	0.043	-0.016	0.063
	(0.03)	(0.08)	(0.09)	(0.05)	(0.09)	(0.07)	(0.10)	(0.10)	(0.11)
TG6 = Statistics+RM LH	0.055	-0.089	-0.106	-0.023	-0.084	-0.13	-0.02	-0.179	0.159
	(0.04)	(0.08)	(0.08)	(0.05)	(0.08)	(0.072)+	(0.09)	(0.089)*	(0.10)
TG7 = Statistics+RM HH	0.124	0.054	0.029	0.009	0.077	0.051	0.196	-0.071	0.269
	(0.050)*	(0.08)	(0.08)	(0.05)	(0.09)	(0.08)	(0.12)	(0.10)	(0.141)+
Constant	0.15	0.68	0.60	85.59	0.68	0.61	0.63	0.78	0.82
	(0.026)**	(0.061)**	(0.063)**	(4.687)**	(0.060)**	(0.052)**	(0.068)**	(0.078)**	(0.074)**
Observations	61380	38262	39007	35990	18487	18774	9085	9261	18346
F-stat (TG3 + 6 = TG4 + 7)	1.58	0.92	0.94	0.49	1.80	2.53	1.80	0.65	0.53

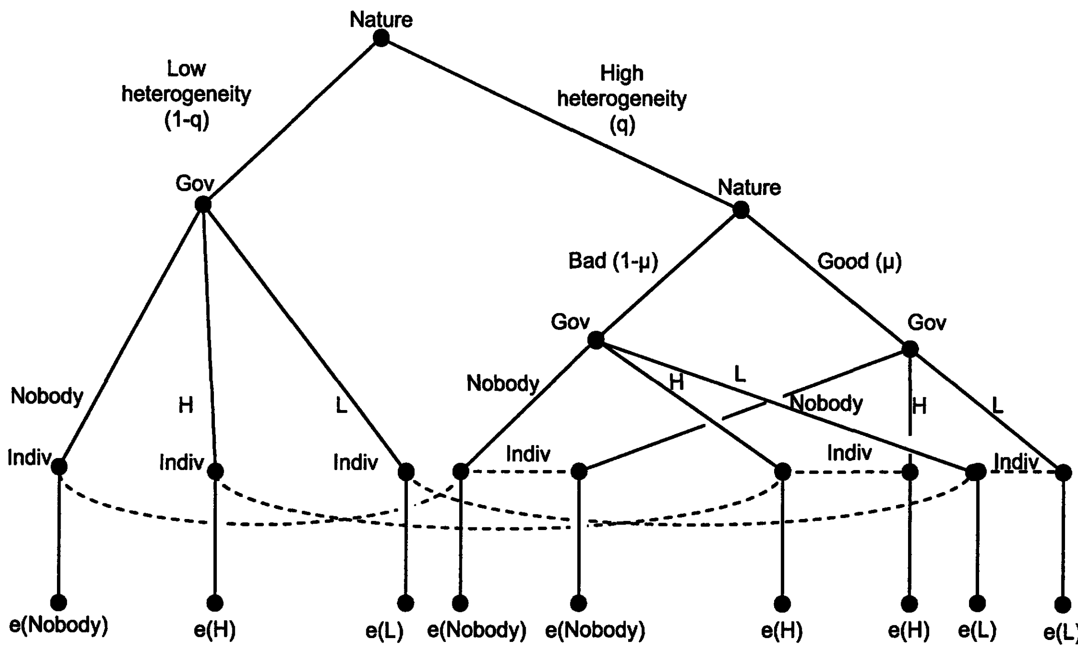
Notes: (#) Gap is the absolute distance |Perceived Returns – Estimated Average Returns|
Omitted categories: Control group. All regressions include dummies for education levels; coefficients not shown.
Standard errors clustered at the school level (621 schools); observations weighted by sampling probability
Columns 5-9: Sample matched to baseline perceived returns data; these regressions control for gap at baseline
Columns 7-9 present results for gap of perceived returns for self
+ significant at 10%; * significant at 5%; ** significant at 1%

Table 7: Impact on Schooling Effort and Test Scores

	Dependent variables								
			(Normalized) Posttest - Pretest (Matched Sample)						
	Attendance	Test Scores	Test Scores	Under-estimated(#)	Over-estimated	Difference (Underestimated – Overestimated)	Rich	Poor	Difference (Rich – Poor)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A									
AnyStatistics	7.846	0.202	0.237	0.365	-0.223	0.576	0.249	0.264	-0.024
	(4.87)	(0.106)+	(0.111)*	(0.156)*	(0.17)	(0.193)**	(0.119)*	(0.137)+	(0.13)
AnyRM	4.947	0.074	0.079	0.112	-0.203	0.32	0.087	0.101	-0.016
	(5.00)	(0.08)	(0.08)	(0.10)	(0.12)	(0.131)*	(0.10)	(0.09)	(0.10)
Statistics+RM	-5.032	-0.132	-0.197	-0.384	0.273	-0.651	-0.217	-0.234	0.022
	(5.36)	(0.13)	(0.13)	(0.178)*	(0.20)	(0.229)**	(0.15)	(0.16)	(0.16)
Panel B									
AnyStatistics	3.517	0.105	0.091	0.079	-0.035	0.109	0.087	0.092	-0.01
	(2.021)+	(0.057)+	(0.06)	(0.08)	(0.11)	(0.12)	(0.08)	(0.07)	(0.09)
Panel C									
AnyRM	0.474	0.01	-0.014	-0.095	-0.105	0.016	-0.018	-0.01	-0.008
	(2.11)	(0.07)	(0.07)	(0.11)	(0.11)	(0.14)	(0.08)	(0.09)	(0.09)
Panel D									
TG1 = Statistics	7.796	0.202	0.237	0.365	-0.223	0.576	0.247	0.264	-0.025
	(4.93)	(0.106)+	(0.111)*	(0.157)*	(0.17)	(0.194)**	(0.119)*	(0.137)+	(0.13)
TG2 = RM LM	7.551	0.088	0.091	0.123	-0.231	0.353	0.164	0.055	0.105
	(5.44)	(0.10)	(0.11)	(0.12)	(0.17)	(0.158)*	(0.13)	(0.13)	(0.13)
TG3 = RM LH	-0.771	0.17	0.156	0.15	-0.123	0.28	0.086	0.271	-0.182
	(6.80)	(0.096)+	(0.090)+	(0.10)	(0.16)	(0.157)+	(0.09)	(0.112)*	(0.104)+
TG4 = RM HH	5.646	-0.022	-0.002	0.045	-0.271	0.326	0.01	0.019	-0.014
	(5.36)	(0.10)	(0.12)	(0.14)	(0.159)+	(0.176)+	(0.16)	(0.11)	(0.17)
TG5 = Statistics+RM LM	5.891	0.117	0.121	0.127	-0.106	0.226	0.009	0.233	-0.242
	(5.11)	(0.11)	(0.11)	(0.12)	(0.13)	(0.15)	(0.15)	(0.126)+	(0.17)
TG6 = Statistics+RM LH	7.098	0.139	0.113	0.077	-0.201	0.282	0.184	0.037	0.15
	(5.30)	(0.11)	(0.12)	(0.17)	(0.24)	(0.27)	(0.15)	(0.12)	(0.12)
TG7 = Statistics+RM HH	9.334	0.182	0.124	0.074	-0.139	0.218	0.138	0.129	0.004
	(4.779)+	(0.106)+	(0.10)	(0.11)	(0.17)	(0.17)	(0.11)	(0.13)	(0.11)
Constant	85.59	-1.24	-1.17	-1.16	-0.93	-0.93	-1.15	-1.20	-1.19
	(4.687)**	(0.066)**	(0.063)**	(0.079)**	(0.102)**	(0.101)**	(0.071)**	(0.074)**	(0.073)**
Observations	176	11659	6297	2877	492	3369	2911	3243	6154

Notes: (#) "Underestimated" denotes the subsample of individuals whose initial perceived returns were below the estimated average returns for some level of education
Omitted categories: Control group.
Column 1: school-level data, all attendance regressions control for school visit time.
Robust standard errors in parentheses; observations weighted by sampling probability and class size.
Columns 2-9: individual-level data. Standard errors clustered at the school level; observations weighted by sampling probability
Test score regressions all control for baseline pretest scores. Posttest scores are normalized by substracting the mean.
and divided by the standard deviation of the control group pretest.
Columns 3-9: Sample matched to baseline perceived returns data.
+ significant at 10%; * significant at 5%; ** significant at 1%

Appendix Figure 1: Extensive-Form Representation of the Game

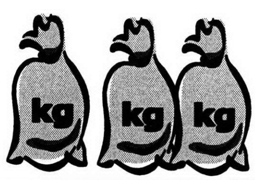
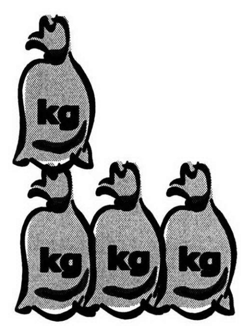
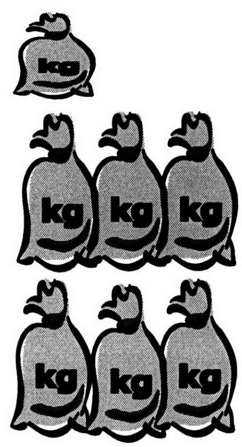
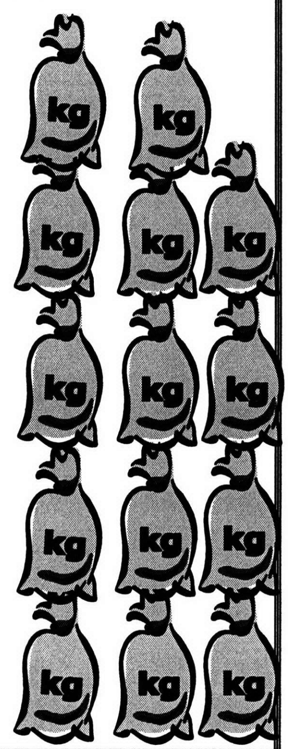


Appendix Figure 2: Information Card

Ministry of Education

RETURNS TO EDUCATION IN MADAGASCAR 2005

This table presents the average monthly income (in ARIARY) of 25 year-old Malagasies, by gender and by different educational levels.

	Without any degree	Primary school with CEPE	Lower secondary school with BEPC	At least high school with BAC
Female	**34 524 Ar**	**44 119 Ar**	**73 771 Ar**	**163 344 Ar**
Male	**47 637 Ar**	**60 877 Ar**	**101 793 Ar**	**225 389 Ar**
Gain				

Source :Calculations from the 2005 Household Survey

Notes: The columns represent education levels (CEPE, BEPC, and BAC are the corresponding degrees); the rows give the mean earnings from household survey. 1 USD = 2000 Ar. The gain from education is illustrated by increased numbers of rice bags.

Chapter 2

Improving Management in Education: Evidence from a Randomized Experiment in Madagascar

2.1 Introduction

Inadequate quality in education is a concern in many developing countries, despite increased funding and inputs that go into the system. In Madagascar, recent reforms since 2002 to provide free primary education as well as free inputs (school kits and basic textbooks) are still insufficient to ensure a basic education. Only 63% of Grade 5 children pass the primary-cycle exam, a minimum level of language and math knowledge presumed at this grade (Tan 2005). The political economy of public good provision argues that government bureaucrats might have better incentives to improve quantity, rather than quality, of services since the benefits of improving service quality are diffuse and harder to verify (Glewwe and Kremer 2005). The World Development Report 2004 thus points to the need of promoting local accountability, that is, enabling the beneficiaries of education to take control of service quality (World Bank 2004).

It remains an open empirical question which interventions, solely top-down approaches or with beneficiary participation, actually improve public services. A top-down approach could aid bureaucrats in their work to monitor school quality. Its impact is based on the premise that these administrators have the right incentives and would carry out effective monitoring. While Olken (2007) shows that external auditors were successful in monitoring corruption and ensuring quality in Indonesian road projects, we cannot say anything definitive about relying on the existing management system to do the same. On the other hand, promoting local accountability may work well if we know exactly what makes beneficiary participation effective. Beneficiary participation takes several forms: parents can monitor teachers either directly via a punishment and reward scheme or indirectly by complaining to the district supervisors. However, participatory programs that require joint actions by a group of parents may face coordination failures and free-riding problems.

To contribute to this debate, we present in this paper evidence from a large-scale randomized experiment in Madagascar's public primary school system. We study the impact of top-down interventions within the existing education system, alone versus in combination with local accountability. The goal of the interventions evaluated here (called AGEMAD[1] program) was to facilitate the administrators' tasks in supervising their schools, on one level. On another level, the motivation was also to provide the village community with information about their school's performance and to allow them to coordinate on taking actions directly. A presumption behind AGEMAD is that lack of information and complementary inputs hold back effective control by the beneficiaries and quality service delivery by the education workers.

Our strategy involves three interventions implemented by the Ministry of Education. The first intervention provided the district administrators with operational tools for their tasks and a training on those tools. The second intervention did this, and in addition, trained and provided respective tools to the subdistrict head. Some examples of the administrators' tools are supervision forms for school visits and information on the performance and resource level at each of their schools. These operational tools are designed to help administrators better supervise the quality of education delivered at the school and allocate resources efficiently within their jurisdiction.

The third intervention had an additional treatment at some randomly selected schools to involve the parents in monitoring school activities. This school-level intervention entailed school meetings organized by the AGEMAD team to provide the parents with the school report card and allowed them to coordinate on doing something about education at the village. Even when the beneficiaries can exercise control and teachers exert more efforts, they may still lack the complementary inputs to improve teaching quality. AGEMAD therefore provided pedagogical and administrative tools

[1] AGEMAD is the French acronym for improving management in education in Madagascar.

for teachers. Examples of such complementary inputs are structured tools for lesson planning, records of student attendance and learning, and reports to parents and to the school director.[2] The other untreated schools in treated districts and treated subdistricts benefited from the top-down interventions but did not receive the report cards and teacher tools directly. It was the responsibility of the district and subdistrict heads to distribute them and implement a monitoring system in those schools.

We collected data on a variety of teaching practices through a school survey. We also collected student attendance data during unannounced visits and student test scores from an achievement test administered independently. We find that the interventions targeted at the district and subdistrict administrators have minimal effects on their behaviors and on the schools and students under their responsibility. This is not because the tools failed entirely: each tool is used by 90% of the subdistrict heads and more than half of the district heads on average. While treated subdistrict heads visited their schools more often, they did not actively involve the parents and the local community in monitoring the school. The teachers in these two treatment groups do not plan their lessons or evaluate their students more, and the students perform no better than those in the comparison group.

On the contrary, the third intervention, implemented at the school level, led to an improvement in several teaching behaviors and student outcomes. Teachers' lesson planning practices improved by 0.26 standard deviations on average, and student evaluation practices by 0.14 standard deviations. Meanwhile, teacher attendance and the school's communication with parents did not improve. This result is consistent with the predictions of a multi-tasking principal-agent model in which teacher efforts are directed toward activities closely tied with observable indicators of their performance, such as those in the book records. The net effect on student attendance was positive:

[2]Evidence to date (Glewwe et. al (2002), Glewwe et. al (2004), and Banerjee et. al (2002)) shows that more resources alone, such as textbooks, flip charts, additional teachers, do not lead to improved learning at the school. Our conjecture is that resources directly aimed at changing the way education is delivered, such as the AGEMAD tools, can work.

4.3 percentage point increase from 87% mean attendance. Test scores of students in this treatment group are 0.1 standard deviations higher than those in the comparison group, as well as those affected by the interventions at higher levels, two years after the implementation of the program.

We interpret these results as suggesting that in the Madagascar context, it is difficult to rely on the existing hierarchy to implement effective monitoring. In the literature evaluating top-down programs, the results presented here differ from Olken (2007). He finds that top-down monitoring reduces corruption by more than the grassroots programs in his experiments. While monitoring in that case was carried out by an external audit agency, the evidence from Madagascar indicates that targeting bureaucrats within the system did not lead to more successful monitoring, especially when our top-down interventions did not introduce explicit incentives for the bureaucrats to do so. In this way, our paper resonates earlier evidence that local government bureaucrats by themselves do not succeed in implementing incentive programs in service delivery (Kremer and Chen (2002), Banerjee, Duflo, and Glennerster (2008)).

The grassroots approach with beneficiary control worked better than the top-down approach in our context. Comparing the results here with the literature, we argue that it was important to arrive at a specific action plan in designing this participatory program. The existing set of field experiments to evaluate the impact of accountability in developing countries gives mixed evidence. Some participatory programs do not work while others do, with specific steps for the beneficiaries to take actions. Kremer and Vermeersch (2005) does not find any improvement in test scores or teacher attendance when parent associations in Kenya are allowed to choose and implement a reward program for teachers. Banerjee et. al. (2008) finds no effect of participatory programs in India, which organized a village meeting to inform the village community about the poor status of learning there and about the control power of the village education committee. The school-level intervention in Madagascar is quite similar to this information project in India with respect to the parent meeting and the school report card components. One difference is that AGEMAD suggested in the school meeting's agenda what the parents can do, aside from the contextual distinction. The

AGEMAD meetings all yielded in an action plan with specific goals and tasks for the school, as well as the actors in charge of doing and monitoring each task. This activity at least ensures no coordination failure at the early stage for the large group of parents to agree on what to do. In this sense, our results are similar to Bjorkman and Svensson (2006), which finds that accountability improves health services in Uganda. The community meetings there to discuss the status of public health services were structured to lead to an action plan. Duflo et. al (2007) also finds support for the use of parental supervision in monitoring teacher practices and improving children's learning. They study an intervention in Kenya that empowered school committees in randomly selected schools with training in how to monitor teachers, and with a formal committee meeting to evaluate teachers' performance. They find that teachers in those schools attend class and teach more often, and their students have higher test scores than those in the schools without empowered parental involvement. In both of these latter studies as well as ours, the parents had specific steps toward to taking actions.

The organization of this paper is as follows. Section 2.2 provides a sketch of the institutional context in Madagascar. Section 2.3 describes the package of interventions and the evaluation design. In section 2.4, we discuss data collection and experimental validity. Section 2.5 presents the empirical strategy and results of the experiment. The final section concludes.

2.2 Primary Education System in Madagascar

On the supply side, the administrative structure of education in Madagascar is a typical hierarchy from the Ministry of Education (MENRS) to districts, subdistricts, and schools. Overall, roughly 2.7 million students are enrolled in 15,000 public primary schools. These schools are grouped into subdistrict zones called ZAPs whereby the ZAP head provides frontline administrative and pedagogical support. These officers report to one of 111 school district (CISCO) heads, who in turn report to MENRS. The CISCO and ZAP heads manage resource flows to schools, oversee teaching and learning practices as well as the collection of school statistics. They also administer the annual national examination at the end of Grade 5 (CEPE exam). Some specific examples of their work include visiting schools for supervision, disbursing teacher paychecks, processing their in-service training and transfer requests, organizing school building and maintenance projects, and distributing books and school grants. The selection of CISCO heads is now nominally competitive, with performance-based reviews after each three-year contract. ZAP heads are usually retired teachers; they rarely face any firing threat.

On the beneficiary side, the main channel of community involvement in primary education in Madagascar has been mostly informal through local parents' associations (FRAM). Recognizing the informal existence of the FRAMs, the government of Madagascar created formal school boards (FAF) starting in 2002-2003. These committees are responsible exclusively for managing capitation grants and not focused on education quality. School boards consist of the village head, the school director, the subdistrict head (rarely there), teacher representatives and the president of FRAM. Together, the FRAM and the school board could exercise control power to pressure the teachers to deliver quality services.

In theory, FRAMs have both direct (social pressure, hire and fire contract teachers[3]) and indirect control (report school problems or complain to higher authorities). In practice, each FRAM is usually a small group consisted of a few nominated parents. First, to influence the school may require joint actions from all the parents rather than just this small group. Coordination failures and freeriding problems may lead to limited participatory actions from the existing FRAMs. Second, they seem to know little about how children at the school are doing, and thus not doing enough to influence the school's operations. Given the possible lack of knowledge of the villagers about the school's performance, providing them with school report cards and organizing local accountability meetings to help parents coordinate on doing something about education might in turn improve

[3]These contract teachers usually hold only a lower secondary or high school degree, hired directly by FRAM and paid on average half the salary of the regular teachers.

incentives for school teachers.

But even if the community puts pressure on the teachers or complains to district and subdistrict heads, these actors themselves do not always have the proper means to actually do their work better. Our observations from the field suggest that teachers do not dispose of basic and standardized tools for their routine tasks to ensure a quality education. Higher-level administrators do not have the right information and tools to monitor, allocate resources, and provide support optimally across the schools under them. It is then possible to improve the quality of education by providing actors at all administrative levels with operational tools to improve their productivity.

2.3 The AGEMAD Interventions

2.3.1 Description and Structure of the Program

The AGEMAD program evaluated in this study was implemented by the Madagascar Ministry of Education (MENRS), with support from the World Bank and the French Development Agency. It is one of MENRS's strategic policies to develop efficient management within the primary education system, with the first 2 school years from September 2005 to June 2007[4] as a pilot study for evaluation. We will refer to 2005-6 as "year 1" and 2006-7 as "year 2."

In 2005, MENRS initiated its AGEMAD program in 3774 public primary schools of 30 rural CISCOs. This purposive sample represents all geographical areas of the country, but focuses on those CISCOs with relatively high repetition rates. For the purpose of the experiment, we excluded schools that are too far away and extremely difficult to access.

As we investigate whether top-down interventions work to improve public services or beneficiary monitoring is also necessary, the AGEMAD initiative took place at three levels in the administrative system: CISCO, ZAP, and school. To give an overview, the CISCO and ZAP interventions provided operational tools for administrators to monitor and support their schools. Meanwhile, the school-

[4]The school year in Madagascar runs from September to June.

level intervention has a "bottom-up" component on top of these treatments, giving parents a way to act directly. First, 15 CISCOs were randomly chosen to receive the CISCO-level treatment; no contact was made in the remaining 15 control CISCOs. The CISCO team received a training and sufficient operational tools to facilitate their routine tasks during the school year. Tools for CISCO heads include forms for supervision visits to the school or procurement sheets for school supplies and school grants. See Appendix Table 4 for the full list of CISCO tools. These are printed templates for the users to fill in according to the attached instructions. One exception is the CISCO report card that each treated CISCO received, which actually has information on the performance (dropout rate, CEPE exam pass rate, and repetition rate) and resource level of all the ZAPs under this CISCO. The CISCO head can find this tool useful in evaluating and monitoring subdistricts.

Second, some ZAPs were selected randomly within treated CISCOs to receive the ZAP-level intervention, which is quite similar to that at the CISCO but with ZAP tools and ZAP report cards. Examples of tools for ZAP heads include forms for supervision visits to the school, community meeting forms, or procurement sheets for school supplies and school grants. The full list is displayed in Appendix Table 3. The ZAP report card has information on the performance and resource level of all the schools under this ZAP. The CISCO and ZAP interventions are two levels of "top-down" programs to facilitate administrators' work in monitoring schools under their jurisdiction. Tool distribution was accompanied by user trainings in September of each year. During training sessions, MENRS representatives emphasized on the core responsibilities of each actor, and explained how to use the tools in performing these responsibilities.

Third, within treated ZAPs, some schools were randomly chosen to receive the school treatment, in addition to the interventions at higher administration. The school-level intervention aimed to improve accountability by sharing information and mobilizing parents to coordinate on an action plan. In particular, MENRS distributed directly to the treated schools their school report card with performance indicators from the previous academic year, and organized accountability meetings for the parents to discuss it. The three simple indicators presented on the report card should be comprehensible to illiterate parents: dropout rate, CEPE exam pass rate, and repetition rate. All statistics were calculated using data from the national school census. Based on these statistics, MENRS classified schools into four categories of performance relative to resource endowment (resource is proxied by the student-teacher ratio). "Exceptional" represents schools with relatively few resources but perform well in terms of dropout rate, CEPE pass rate, or repetition rate. "Satisfactory" implies schools with good resources and good performance. "Difficulty" means having few resources and performing poorly. "Disappointing" schools are those with good resources but performing poorly. A school may be satisfactory in terms of exam pass rate but disappointing in terms of repetition rate. An example of the school report card (translated from the original version in French and Malagasy) is included in Appendix Figure 1.

To help parents focus attention and action on results as shown in the report card, AGEMAD also organized two accountability meetings between the treated school and the parents as well as the local community. One took place at the beginning and the other near the end of the school year. MENRS suggested an agenda for these half-day school meetings and requested that an action plan was set up at the first meeting of the year and evaluated at the last meeting, but the actual content of discussion was not mandated and left to the discretion of the participants.[5] Proposed items on the agenda include a review of the current situation at the school, its progress with information available from the report card, discussion on possible solutions to existing problems, and setting goals in an actual action plan. To arrive at a specific plan, the participants first identified the issues hampering academic achievements and then prioritized the activities to overcome these problems. The action plan had to be approved by the general audience at the parent meeting. While the level of details varies from school to school, it typically has a list of objectives or desired improvement in certain general activities such as teaching. For each objective, several tasks were identified to be

[5] MENRS did not send any of its staff to the school meetings. About half of the times, the ZAP head is present, but that is part of his/her job and not really the presence of a new authority figure. MENRS had also hired facilitators to ensure that school meetings indeed took place and to collect information on the meetings. One facilitator would attend each meeting, but would usually not intervene.

performed over a specified time period, along with the actors responsible for them. One example of the goals is to increase the exam pass rate at the school by 5 percentage points by the end of the academic year. The common tasks specified for the teachers include lesson planning and student evaluation every few weeks. Each parent and the parent association are expected to monitor the student evaluation reports that the teachers are supposed to communicate to them.

These treated schools also received director and teacher tools directly. Teacher tools are, for example, lesson-planning notebooks, class attendance registers, and grade registers. The director would have the registration roster, summary of student test scores, and community meeting forms. Appendix Tables 1-2 list all these tools, designed to be user-friendly and time-efficient. They were produced as physical sheets of paper or notebooks, for the users to fill in throughout the school year. The presumed benefit of the tools is to provide inputs complementary to education workers' effort in their work, in turn improving school quality.

The other schools in the treated ZAPs and 15 treated CISCOs did not receive any direct school-level intervention. Report cards and sample teacher tools were created for them, but it was left to the hands of the CISCO and ZAP heads to distribute, implement the tools, and monitor school activities. There was no option to have a treatment group with school report cards alone since it would have been practically difficult to work with government schools in Madagascar without going through the district and subdistrict heads.

The treatment design with the relevant samples of observations is demonstrated in Table 1. Treatment assignment was random, stratified by baseline repetition rate and school size (and geographical region for the CISCO randomization). The above levels of treatment intensity generate four types of schools. The control group includes all the schools in the 15 control CISCOs. The first treatment group refers to those schools influenced by the program only at their CISCO (schools in AGEMAD CISCOs). The second group of treated schools is influenced by interventions at both the CISCO and ZAP levels (schools in AGEMAD ZAPs). These two treatment groups represent a top-down approach to improving school quality. It is the responsibility of the CISCO and ZAP administrators to monitor the schools. The third type of treatment schools (called AGEMAD schools) received all three levels of interventions, CISCO/ZAP/school. The additional bottom-up component here engaged parents into acting directly rather than depending on the bureaucrats. Due to budget constraints, we randomly selected 303 schools in each of the four groups, i.e. 1212 schools total, as the follow-up sample for data collection.[6]

This design allows us to learn which approach in the AGEMAD interventions effectively improve the way schools function and student performance. The top-down approach may work if the CISCO-level and ZAP-level interventions produced sufficient impact on service provider behaviors and on key schooling outcomes. On the other hand, such interventions might be ineffective unless reinforced by direct involvement of the parents through accountability meetings at the school.

Our tracking data suggests that the interventions were well implemented with good take-up. All the AGEMAD schools that were intended actually received the tools and organized the meetings. Accountability meetings were well attended with 64 participants on average, half of whom were women. 30% of the attendees spoke up and asked questions at least once. 94% of these accountability meetings discussed the school report cards and students' learning results as intended. On average, 85% of the teachers in AGEMAD schools reported using several tools, with the pedagogical and evaluation tools used by almost everyone. Tool usage rate is lower for the director (65%). Of those who used the tools, close to 100% found them useful.

2.3.2 Why Might We Expect These Interventions to Work?

Teacher effort is an important input in students' learning, so higher teacher effort would improve the quality of education at school. In theory, we would expect that most means of monitoring could potentially give teachers better incentives to perform their duty, thus leading to higher test score achievement for students.

[6] Power calculations showed that this sample size is sufficient to detect an increase of 0.25 standard deviations in school-level average test scores with 95% confidence. Since we could later collect individual-level test scores, this sample of schools should allow us to detect a smaller treatment effect.

The top-down interventions need not always produce an impact. If the CISCO and ZAP administrators in treated districts actually monitor school activities, well-audited schools may perform better. The question is whether we can rely on the existing hierarchy to carry out effective monitoring. These interventions were implemented within the current institutions, almost entirely by the Ministry rather than a substitute system of management. The ZAP and CISCO heads do not face direct reward and punishment mechanisms for monitoring their schools well. It is ambiguous if the ZAP and CISCO interventions bring about strong changes in behaviors and student outcomes.

Alternatively, the school intervention provides information on school performance and allows the parents to coordinate on taking actions to monitor service quality. School report cards were rare in rural Madagascar, and centrally-generated report cards were non-existent before the program. This component of AGEMAD therefore provided probably new information on the quality of current services. With that, parents can exercise social pressure on the teachers or implement explicit incentive schemes so that the school does a better job of teaching. The school-level treatment is expected to affect teacher behaviors if the parents register the information provided on the report card, believe that they can exercise their control power, and successfully coordinate. This is likely the case in our intervention. Most parents attended the school meetings and participated actively in the discussion of the report card and of an action plan.

2.4 Evaluation

2.4.1 Data Collection

We expect the interventions to affect intermediate outcomes such as behaviors of education workers and the functioning of the school. This data comes mainly from the school questionnaire in 1212 schools of the follow-up sample. During each of the two school years between January and May, MENRS hired an NGO (Aide et Action) to administer a school questionnaire during randomly timed, unannounced visits to the schools in the follow-up sample. In addition to spot-checking teacher and student absenteeism, the surveyor interviewed the school director and teachers about their usage of common tools and their administrative and pedagogical practices. For example, some self-reported measures include: pedagogical organization, student assessment and attendance registration, and communication with the parents and local school community. The year 1 questionnaire took place right after the interventions while teachers were just familiarizing themselves with the tools. Since we have better data in year 2, we will report the results from this latter school questionnaire.

In year 2, we also surveyed the ZAP heads on their interactions with their schools. In particular, we will see if treated ZAP heads pay more visits and attend more meetings at schools in difficulty as notified in the report card.

The next outcome of interest, student attendance, was collected by surveyors during unannounced school visits. From this data, we constructed three measures of attendance rates. First, "classroom count" is the ratio of student attendance in each classroom to its enrollment. This number is reported during individual teacher interviews with the surveyor. Second, "surveyor reported" is the ratio of how many students in the whole school the surveyor can quickly see upon arrival, divided by the enrollment of classes in session. This data is not the most precise since the surveyor did not have enough time to do a headcount, so he or she would give a rough guess instead. Third, "director's book record" is attendance rate from the school director's official record. Dropout and repetition rates are also available in the year-end statistical form (*Fiche de Fin d'Année)* administered nationally by MENRS.

We are interested in whether and what levels of interventions ultimately improve learning. Our primary data on learning comes from achievement test scores collected in both years of the experiment. MENRS administered the year 1 test to Grade 3 students in February 2006 (close to baseline). The endline test was administered to the same children in May 2007 (then in Grade 4). The testing instrument was based on PASEC tests[7] and already well-developed with effective testing

[7]PASEC (Programme d'Analyse des Systèmes Educatifs de la CONFEMEN) is a program in 15 francophone

procedures. These two tests are intended to measure children's competency in three materials: math, French, and Malagasy. They include basic calculations and grammar questions at the level that Grade 3 and Grade 4 students, respectively, are supposed to master. We normalize the endline test scores by subtracting the mean, then dividing by the standard deviation of the control group.

Moreover, the national school census serves as a secondary source of learning data. CEPE (end-of-primary-cycle national examination) exam results for Grade 5 students are available at the school level.

2.4.2 Descriptive Statistics

Table 2 presents descriptive statistics for a range of school characteristics prior to the interventions, and also confirms that the treatment and control groups do not differ systematically, as expected given the random assignment. The average school has more than 200 students, with roughly 54 students per teacher. 66% of Grade 5 students in these schools pass the minimum learning requirement, the CEPE exam. 18% of students repeat a grade, and 6% of students enrolled at the beginning of the school year drop out by the end. Columns 2-4 show that the treatment groups are statistically similar to the comparison group in terms of these characteristics.

2.4.3 Attrition

To minimize the potential bias caused by differential attrition, we tried to measure outcome variables for all the original participants of the program. The endline test was administered to as many of the baseline children as possible. The school director had asked children who no longer attended school to come at the day of the test; home visits were made to a random subset of the attriters.

Table 3 shows attrition rates by treatment assignment. The first row suggests that attrition is roughly 13% and is similar across treatment status. A further check is on the type of children who have attrited from our sample. If we look at their baseline test scores and their gender, differences

countries that studies elements of learning for students.

between the attriters and stayers are small and again similar in our treatment and control groups.

2.5 Estimation Strategy and Results

We first discuss the interventions' impact on intermediate outcomes such as teacher behaviors, director behaviors and general functioning of the school, as well as behaviors of the ZAP and CISCO heads. Then we report treatment effects on student attendance, dropout rate, and learning.

The main estimating equation used throughout this analysis is

$$Y_{is} = \alpha + \beta_1 * \mathrm{TG1}_s + \beta_2 * \mathrm{TG2}_s + \beta_3 * \mathrm{TG3}_s + \delta X_{is} + \varepsilon_{is} \tag{2.1}$$

where Y_{is} represents outcome variables, on behaviors and on students' performance, of individual i in school s. *TG's* denote treatment group indicators, with the control group being the omitted category. TG1 is an indicator for whether the school is in a treated CISCO. TG2 is an indicator for whether the school is in a treated ZAP. TG3 is an indicator for whether the school itself received the AGEMAD interventions. Given the randomization, our coefficients of interest β's are interpretable as the causal impact of different levels of the AGEMAD treatment. We control for various baseline school or individual characteristics X. All the regressions using the school questionnaire data control for the month of random school visits to account for seasonal variation.

We also run an alternative regression with a dummy for any treatment at all (1, 2, or 3) in addition to the treatment 3 dummy. We can interpret β as the effect of Treatments 1 or 2 versus the control group, and β_4 as the effect of Treatment 3 versus Treatments 1 or 2.

$$Y_{is} = \alpha + \beta * \mathrm{AnyTreatment}_s + \beta_4 * \mathrm{TG3}_s + \delta X_{is} + \varepsilon_{is} \tag{2.2}$$

Both equations are estimated using OLS, with standard errors clustered at the CISCO level since treatments were implemented at this level.[8] Observations are weighted by the probability of selection.

Since there are many intermediate outcomes on behaviors that may have been affected by AGEMAD, we adopt the approach from Katz, Kling, and Liebman (2007) when appropriate. We group outcomes into categories of related variables and report the average effect of each category: lesson planning, student evaluation practices, communication and meetings to discuss student matters, and so on. For example, the average standardized effect of Treatment 3 for a category of K outcomes, each indexed by k, is defined as $\beta_3 = \frac{1}{K}\sum_{k=1}^{K}\frac{\beta_{3k}}{\sigma_k}$ where β_{3k} is the coefficient from equation (2.1) for outcome k, and σ_k is the standard deviation of this outcome. We calculate the standard errors of this average effect by estimating a system of seemingly unrelated regressions (SUR) for the outcomes in each category.

2.5.1 Intermediate Outcomes at School

2.5.1.1 Impact of CISCO and ZAP Interventions on Intermediate Outcomes

The first set of intermediate outcomes at the school is teacher behaviors, as measured in the school survey. Table 4 reports treatment effects on teacher attendance and pedagogical practices. The first set of *Regression (1)* reports OLS results of equation (2.1) for various outcome variables of teacher behaviors, listed on the leftmost column. The second set of *Regression (2)* runs equation (2.2) and gives quite similar results with more precise estimates. These results are robust to a varying set of teacher and school control variables.

We find that schools benefiting from the CISCO and ZAP interventions do not generally have better teacher practices. The SUR analysis for lesson planning is presented in Panel B. Columns 3, 4, and 6 show no significant improvement in the schools that received the district and subdistrict interventions. Teachers of these schools (treatment groups 1 and 2) are on average 0.1 standard

[8] A probit estimation instead of the linear probability version of equations (2.1) and (2.2) provides very similar results. The estimates are also similar if we restrict the sample to treatment groups 2 and 3 and cluster the standard errors at the school level.

deviations more likely to perform these tasks, but the estimate is statistically indistinguishable from zero. Similarly in terms of student evaluation practices, column 6 in Panel C suggests that schools in the treated districts and sub-districts do not perform these tasks better than the control schools. We will return to this point when we discuss the ZAP's behaviors.

As shown in Panel A of Table 4, the CISCO and ZAP interventions do not improve teacher attendance. We study three measures of teacher attendance. That reported by the director during the random check is attendance at the teacher level, while the other two measures are at the school level. The third measure, percentage of teachers present reported by the surveyor during his random visit, comes from an approximate headcount upon arrival and is relatively noisy. The mean attendance rate in column 1 suggests less than 10% absenteeism, in contrast to higher teacher absenteeism in Kenya–20% (Glewwe et. al. 2003) or India–25% (Chaudhury et. al. 2006). The coefficient estimates of treatment effects for treatments 1 and 2 are often negative in columns 3, 4, and 6, though very small and hardly distinguishable from zero.

The second set of intermediate outcomes at the school is the director's behaviors and school functioning. We collected some observational data at the outset of each unannounced school visit to measure general school functioning, but find no sign of significant improvement in those schools receiving the top-down interventions. According to Panels A and B of Table 5, there is no strong difference between treatment 1 or 2 and the control schools concerning whether students are playing in the hall during class time, whether the teachers are teaching, and the chance that a classroom has a teacher or a lesson in progress. The results on control and reporting practices by the director are displayed in Panel C. Our estimation of regression (2) implies zero average treatment effect for schools in the district and subdistrict interventions versus the control group. The estimate of -0.04 standard deviations is statistically insignificant. Together, one interpretation of these no-impact findings for the top-down interventions is that the treated CISCO and ZAP heads do not by themselves implement more effective monitoring.

2.5.1.2 Impact of School Intervention on Intermediate Outcomes

In contrast to the above results on the CISCO and ZAP interventions, the additional school-level intervention appears to be effective in monitoring some school activities. As displayed in Table 4, teachers' practices are improved in AGEMAD schools along some dimensions. Lesson planning is performed much more frequently in AGEMAD schools than in all the others (Panel B, column 2). From a SUR analysis, we find that teachers of AGEMAD reinforced schools (treatment group 3) are on average 0.26 standard deviations more likely to perform these tasks. This estimate is statistically significant at the 5% level. Student evaluation (Panel C) has improved on average by 0.14 standard deviations in AGEMAD schools compared to control schools. This estimate has large standard errors, and is statistically significant only in regression (2). As implied in column 5, while schools in treated districts and sub-districts do not improve student evaluation practices, schools treated directly with the parents' involvement fare better by 0.07 standard deviations. These results suggest that allowing the beneficiaries to coordinate on exercising their control power to improve education leads to better teaching quality in this context.

However, teacher attendance does not improve in AGEMAD schools even though the director says that (s)he records teacher absences more frequently. The coefficient estimates on treatment 3 in Panel A of Table 4 are often negative, though very small and hardly distinguishable from zero. It may be costly for the parents to verify teacher presence. Then as long as the school director states in the official school records that teachers still show up regularly, there is no additional incentive induced by AGEMAD for teachers to come to school more regularly. Contract teachers are associated with lower presence rate than regular teachers, but the treatments do not affect their attendance differentially.

While teachers say they talk to the director often, only about half of them communicated with some parents over the previous month. Teachers in treated schools are no more likely to communicate with the parents or with the director on student matters, as presented in Panel D of

Table 4. Overall, we do see that teacher behaviors improve on dimensions that are easily recorded for future evaluation of their work (filling out lesson planning and giving exams) and do not improve on other dimensions (teacher attendance and communication with the parents and the director). This finding is consistent with a multi-tasking principal-agent model in which teacher efforts are directed toward activities closely tied with observable evaluation indicators of their performance.

Table 5 presents the impact results of the additional school-level intervention on director behaviors and school functioning. While the standard errors in regression (1) are large, we find in regression (2) an average treatment effect of 0.07 standard deviations better control and reporting practices by the director in AGEMAD schools versus those in the district and subdistrict interventions. This average effect estimate, reported in Panel C column 5, is statistically significant at the 10% level. In particular, the director claims he is 8.5 percentage points more likely to record teacher absenteeism in treated schools (column 2, significant at the 1% level). Nonetheless, as we have discussed, teacher attendance does not improve. Self-reported monitoring from the director is either inaccurate or does not appear effective.

Similar to our findings for the teachers, there is no significant improvement in terms of the director's communication with teachers and with the community on school matters. While the estimates in Panel D are not small (average effects around 0.08 standard deviations), they are imprecise. According to Panels A and B of Table 5, we do not find significant improvement in general school functioning in AGEMAD schools. In sum, the school-level intervention affects mainly the teaching activities at school and the director's monitoring practices.

2.5.2 Behaviors of ZAP and CISCO Heads

To better understand the contrast between the impact of top-down interventions alone versus in combination with parental involvement, this section investigates further some behaviors of the ZAP and CISCO heads. We first look at how the different interventions affect the frequency of their school supervision. Then, we study the effect of the ZAP intervention on how ZAP heads allocate their efforts across supervising different types of schools under them.

We find that all the AGEMAD interventions lead to more school visits by the ZAP head, but (s)he does not hold more meetings with the local community. Specifically, AGEMAD schools and schools in AGEMAD ZAPs receive more frequent visits from the authorities. Panel E of Table 5 presents the impact of the AGEMAD treatments on the number of days since the last visit from the ZAP and CISCO. It is not surprising that CISCO heads tend to visit schools less than ZAP heads. On average, the last visit is 0.32 standard deviations more recent in AGEMAD schools than in comparison schools, and 0.3 standard deviations more recent in schools of treated ZAPs (treatment 2). Despite this evidence of more visits from the authorities, we still find no improvement in pedagogical practices at treatment 2 schools, implying that monitoring from the ZAP head does not appear effective.

In addition, we look at a few self-reported behaviors of ZAP heads from the ZAP questionnaire. The results from this analysis suggest that treatment ZAPs (treatment groups 2 or 3) do not hold more meetings with the local community and the school board. To reconcile this result with the above finding on more ZAP visits from the school questionnaire, one possible interpretation is that treated ZAP heads exercise more school visits but ignore the accountability channel of involving the village community.

In terms of effort allocation by the ZAP head, treated ZAPs were given information on the performance of schools in their jurisdiction. As described in section 2.3.1, the report cards classify schools in each ZAP into four categories: exceptional, in difficulty, satisfactory, and disappointing. We expect treated ZAPs to provide differentially more support to schools in difficulty and exceptional schools than to satisfactory schools if they want to increase overall performance in the ZAP. In contrast to the spirit promoted in AGEMAD report cards, we do not find strong evidence for better allocation of efforts by the ZAP heads toward "schools in difficulty."

Table 6 presents the effects of the ZAP intervention on the number of meetings and visits that each school receives from the ZAP head in year 1, interacted with the school's initial performance category. Schools are categorized by three indicators prior to the program: net exit rate,[9] primary-cycle exam pass rate, and repetition rate. Each column in this table represents a regression. The estimates on the school category terms suggest that exceptional schools and schools in difficulty in the comparison group tend to receive fewer meetings and visits from the ZAP administrator than satisfactory schools. Coefficients on the interaction terms tell us, if anything, the opposite of what we expected although the estimates are statistically insignificant. In general, treated ZAP heads have even relatively fewer interactions with exceptional and difficulty schools than with their satisfactory schools. They did not seem to act based on the information provided to them on their schools' performance.

2.5.3 Student Attendance and Dropouts

With the CISCO and ZAP interventions doing little to affect routine activities at the school, it is not surprising that we find no evidence for their impact on student attendance. Student attendance during the school year was collected by surveyors during unannounced school visits. We have three measures of attendance rates: "classroom count" is at the classroom level, while the other two measures are at the school level. "Surveyor reported" attendance rate is the ratio of the number of students in the whole school that the surveyor can quickly see upon arrival, divided by enrollment of the classes in session. Table 7 reports the treatment effects on attendance, for each of the three measures. As shown in columns 3, 4, and 6, the district and subdistrict intervention dummies always have coefficients statistically indistinguishable from zero, and smaller than that of the school intervention.

The school intervention, on the other hand, led to some changes in behaviors. We have seen that there are better pedagogical practices at AGEMAD schools, so we would expect the treatment

[9]Net exit can be understood as dropout rate. For terminology purposes, the Ministry calls this measure net exit since schools only report enrollment at the beginning and at the end of the year. We only know how many children have left each school on net, but not whether some have transferred or dropped out completely. Student transfers are rare in Madagascar, so net exit from school is almost identical to the dropout rate.

effects to translate into improved student participation at school, including regular attendance and less dropout at the end of the school year, and improved learning. Student attendance rate (in percent) is overall better in AGEMAD schools and not in the other treatments. As shown in Table 7, AGEMAD schools have 4.29 percentage points higher attendance in the director's record compared to the control group average of 87%. Classroom count attendance also has a positive treatment coefficient, but not statistically significant. The regression using the surveyor's estimate of attendance also has insignificant treatment coefficients.

The general observation from this experiment is the lack of impact in the top-down interventions, but the school intervention worked. One exception to this general observation is the result on self-reported dropout rate. Madagascar's administrative data, reported by the schools themselves, reveals positive treatment effects on student performance at the end of year 1, as displayed in Table 8.[10] AGEMAD schools perform on average 0.17 standard deviations higher than the comparison schools, in terms of exam attendance rate, pass rate and dropout. Grade repetition in year 2 is also 5.1 percentage points lower, with both estimates statistically significant. In these regressions, the district and subdistrict interventions have an equally strong positive impact as the school-level intervention. This result diverges from the evidence on the other outcome measures we have thus far. One possible explanation is that self-reported dropout rates are relatively easy to manipulate on the statistical form submitted to the Ministry, as opposed to teachers' lesson planning practices or student test scores measured by a third party. Even if there is no true reduction in dropout rates, schools in the treated districts, possibly feeling they were under extra scrutiny by the Ministry, might have reported better rates.

[10] This table reports OLS results, and they are similar when we run a difference-in-differences regression by including the pre-AGEMAD values of these outcome variables. So we report OLS to be consistent with the rest of the paper.

2.5.4 Learning

To investigate the program effects on the intensive margin of learning, we exploit Grade 4 student test score data from the achievement test at the end of year 2. Table 9 displays regression results of equations (2.1) and (2.2), in which the dependent variable is normalized test score. As shown in columns 2, 3, and 5, treatments solely at the CISCO or ZAP levels both have close to zero impacts, and the estimates are insignificant. This finding is not surprising given that there is no evidence for improvement in teacher behaviors in these two top-down treatment groups.

On the other hand, the magnitudes of regression (1) coefficient estimates suggest that students in the AGEMAD reinforced schools perform 0.095 standard deviations better than those in the control schools. Regression (2) attempts to alleviate the power issue by pooling treatments 1 and 2 together. The coefficient on Treatment 3 (AGEMAD schools) is significant at the 10% level. This result on learning is consistent with the general distinction in treatment effects between the top-down interventions alone and in combination with the school intervention.

This impact of the school intervention on the total score reflects impacts on math and Malagasy subject scores, with the exception of French not improving at all in AGEMAD schools. Math scores improve by 0.12 standard deviations, while Malagasy scores improve by 0.08 standard deviations. But in the Madagascar context, teachers do not master French very well. The AGEMAD interventions, as designed, do not help them teach better something they do not know so well themselves.

2.6 Conclusion

This study contributes to a growing quest to understand better which kinds of top-down or participatory programs might or might not work to improve service quality. We provide evidence from a large randomized experiment in Madagascar's public primary schools. This AGEMAD program had two levels of top-down interventions within the current system, where it distributed operational tools to enable better school monitoring by the district and subdistrict heads. Some randomly selected schools received the school-level treatment on top of the above. As in several other participatory programs, the school intervention here organized accountability meetings between the school and the parents to discuss information on the school report card, and encouraged the establishment of an action plan for the school. We expect the different AGEMAD interventions to influence first-hands the teachers' behaviors if monitoring by the administrators or by the parents is effective, and in turn influence students' learning.

We find that the top-down policy design with mediated control within the system by itself does not seem effective. Providing tools and school information to higher-level bureaucrats did not appear to have much impact on the schools and students under them. While the subdistrict heads claimed that they visited their schools more often, they did not hold more meetings to involve the local community into monitoring the school. There is no evidence of improvement in teaching practices and learning outcomes in these treatment groups.

On the contrary, the interventions down to the school level with beneficiary control had positive impacts. Teacher behaviors improved for some teaching tasks, such as planning lessons and giving evaluation quizzes and tests. Other tasks, less likely to show up in the book records, did not improve: teacher attendance or communication with the parents and the director. On balance, student outcomes still improved in these treated schools. In particular, test scores increased by 0.1 standard deviations after two years. This result suggests that teaching activities (lesson planning and student assessment) are important inputs in influencing student learning.

In sum, the top-down interventions in Madagascar seem ineffective in monitoring the school. While we have seen evidence that external auditors can successfully monitor service delivery, it remains a challenge to depend on bureaucrats in the system to perform that job. In this context, it is the school-level intervention with beneficiary monitoring that worked to hold the school accountable and affected teaching activities and learning.

References

BANERJEE, A., R. BANERJI, E. DUFLO, R. GLENNERSTER AND S. KHEMANI (2008): "Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India," mimeo, MIT.

BANERJEE, A. , E. DUFLO, AND R. GLENNERSTER, (2008): "Putting a Band-Aid on a Corpse: Incentives for Nurses in the Indian Public Health Care System," *Journal of European Economic Association*, Proceedings, 6.2-3, April-May, forthcoming.

BANERJEE, A., S. JACOB AND M. KREMER (2002): "Promoting School Participation in Rural Rajasthan: Results from Some Prospective Trials", mimeo, MIT.

BJÖRKMAN, M. AND J. SVENSSON (2006): "Power to the People: Evidence from a Randomized Experiment of a Citizen Report Card Project in Uganda," mimeo, IIES, Stockholm University.

CHAUDHURY, N., J. HAMMER, M. KREMER, K. MURALIDHARAN AND H. ROGERS (2006): "Missing in Action: Teacher and Health Worker Absence in Developing Countries." *Journal of Economic Perspectives,* 20(1), 91-116.

DUFLO, E., P. DUPAS AND M. KREMER (2007): "Peer Effects, Pupil-Teacher Ratios, and Teacher Incentives: Evidence from a Randomized Evaluation in Kenya." Mimeo, MIT.

GLEWWE, P. AND M. KREMER (2005): "Schools, Teachers, and Education Outcomes in Developing Countries," *Handbook on the Economics of Education* (forthcoming).

GLEWWE, P., M. KREMER AND S. MOULIN (2002): "Textbooks and Test Scores: Evidence from a Prospective Evaluation in Kenya," BREAD Working Paper, Cambridge, MA.

GLEWWE, P., M. KREMER, S. MOULIN AND E. ZITZEWITZ (2004): "Retrospective vs. Prospective Analysis of School Inputs: The Case of Flip Charts in Kenya." *Journal of Development Economics*, 74(1), 251-268.

GLEWWE, P., N. ILIAS AND M. KREMER (2003): "Teacher Incentives," NBER Working Paper 9671.

KREMER, M. AND D. CHEN (2002): "Interim Report on a Teacher Attendance Incentive Program in Kenya", manuscript, Harvard University.

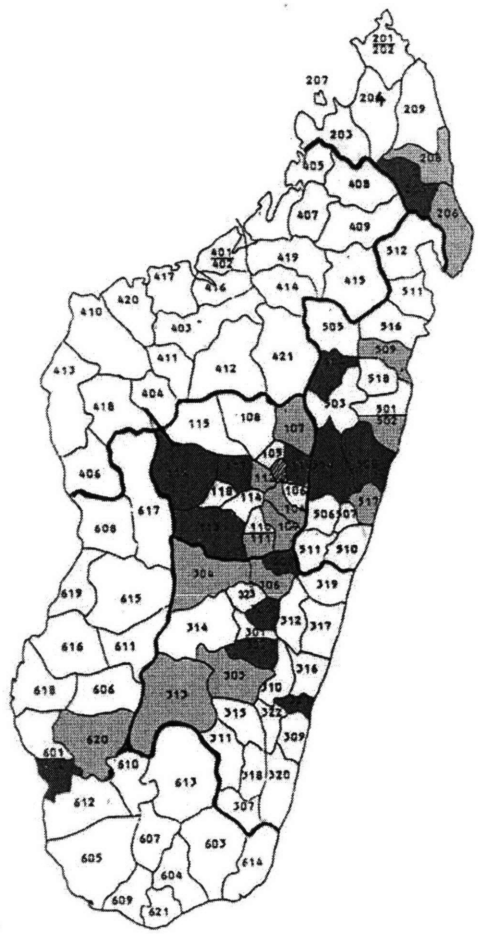
KREMER, M. AND C. VERMEESH (2005): "School Committee Empowerment: Preliminary notes." Mimeo, Harvard University.

OLKEN, B. (2007): "Monitoring Corruption: Evidence from a Field Experiment in Indonesia," *Journal of Political Economy 115 (2)*, pp. 200-249.

TAN, J.P. (2005): "Managing for Results in Primary Education in Madagascar: Evaluating the Impact of Selected Interventions," Concept Note, World Bank.

WORLD BANK (2004): *World Development Report 2004: Making Services Work for the Poor*, New York: Oxford University Press.

Figure 1: School District Map of Madagascar



Notes: The shaded areas are the 30 school districts in our study sample. The darker ones are the treated CISCOs.

Table 1: Experimental Design

	Sample of observations			Description of interventions
	CISCOs	ZAPs	Schools	
Comparison	15	207	1721 (303)	No intervention
Treatment 1	15	170	1314 (303)	CISCO-level intervention
Treatment 2		89	436 (303)	CISCO + ZAP-level interventions
Treatment 3			303 (303)	CISCO + ZAP + School-level interventions

Notes: In parentheses are numbers of schools in the follow-up sample for data collection

Table 2: Baseline School Characteristics

	Comparison Group Mean	Treatment 3-Comparison	Treatment 2-Comparison	Treatment 1-Comparison	Obs.
	(1)	(2)	(3)	(4)	(5)
Panel A: Followup sample					
Enrollment	239.465	-4.785	-25.323	-39.267	1212
		(28.21)	(24.89)	(23.25)	
Repetition Rate	0.177	0.037	0.022	0.029	1212
		(0.02)	(0.02)	(0.02)	
No. of Class Sections	5.693	-0.017	-0.35	-0.366	1212
		(0.30)	(0.25)	(0.24)	
No. of Teachers	4.413	0.033	-0.383	-0.234	1212
		(0.41)	(0.33)	(0.34)	

Panel B: Sample of treated districts

		Treatment 3 Group Mean	Treatment 2-Treatment 3	Treatment 1-Treatment 3	Obs.
	(1)	(2)	(3)	(4)	(5)
Net Exit Rate		5.872	0.931	0.729	873
			(1.10)	(1.11)	
CEPE Exam Pass Rate		66.260	-1.375	-2.572	707
			(3.22)	(3.29)	

Notes: Panel A reports mean differences across groups for all 1212 schools in the followup sample.
Panel B reports mean differences in indicators for which we have baseline data for only the 15 treated CISCOs.
Standard errors in parentheses, clustered at the CISCO level
* significant at 5%; ** significant at 1%

Table 3: Attrition in Endline Test Scores

	Comparison Group Mean	Treatment 3-Comparison	Treatment 2-Comparison	Treatment 1-Comparison
	(1)	(2)	(3)	(4)
Percent Attrition	0.132	0.007	-0.005	0.002
		(0.02)	(0.02)	(0.02)
Difference in pretest	-0.068	-0.143	0.002	-0.012
Attriters - Stayers		(0.08)	(0.10)	(0.08)
Difference in proportion of males	-0.013	0.037	0.011	0.019
Attriters - Stayers		(0.03)	(0.02)	(0.02)

Notes: N=24341

Standard errors in parentheses are clustered at the CISCO level.

Attrition rates in posttest data are at the individual level.

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

Table 4: Teacher Attendance and Practices

	Endline C.Group Mean	*Regression (1)*			*Regression (2)*		
		Treatment 3	Treatment 2	Treatment 1	Treatment 3	Any Treatment	Obs.
Dependent variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Teacher presence							
Director reported during	0.903	-0.03	-0.021	-0.007	-0.015	-0.014	5370
random visit (teacher-level)	(0.008)	(0.030)	(0.019)	(0.021)	(0.022)	(0.019)	
Director's book record	0.939	0.012	0.014	0.008	0.001	0.011	1061
	(0.007)	(0.020)	(0.016)	(0.018)	(0.009)	(0.017)	
Surveyor reported during	0.974	-0.022	-0.017	-0.015	-0.007	-0.016	1069
random visit	(0.005)	(0.018)	(0.012)	(0.010)	(0.014)	(0.009)	
Panel B: Teachers' lesson planning practices							
Prepared a lesson for today	0.798	0.093	0.017	0.05	0.06	0.033	4184
	(0.012)	(0.061)	(0.064)	(0.063)	(0.018)**	(0.062)	
Discussed lesson plan	0.835	0.065	0.014	0.037	0.04	0.025	4184
with director	(0.011)	(0.083)	(0.082)	(0.081)	(0.023)+	(0.080)	
Made weekly lesson plans	0.822	0.104	0.046	0.04	0.061	0.043	4184
	(0.012)	(0.040)**	(0.045)	(0.042)	(0.016)**	(0.043)	
Average Effect (in sd)		**0.260**	**0.077**	**0.124**	**0.160**	**0.100**	**4184**
		(0.117)*	**(0.126)**	**(0.119)**	**(0.046)****	**(0.121)**	
Panel C: Teachers' student evaluation practices							
Evaluated student's learning	0.736	0.105	0.077	0.08	0.027	0.078	4184
last week	(0.014)	(0.140)	(0.137)	(0.138)	(0.017)	(0.137)	
Gave bimonthly exams	0.967	0.004	-0.003	-0.017	0.013	-0.01	4184
	(0.006)	(0.011)	(0.012)	(0.017)	(0.011)	(0.013)	
Average Effect (in sd)		**0.141**	**0.086**	**0.050**	**0.073**	**0.069**	**4184**
		(0.168)	**(0.159)**	**(0.165)**	**(0.042)+**	**(0.161)**	
Panel D: Teachers' communication							
Communicated with parents	0.943	0.017	-0.003	0.013	0.012	0.005	4184
about student performance	(0.007)	(0.027)	(0.029)	(0.024)	(0.02)	(0.023)	
Communicated with parents	0.580	-0.056	-0.085	-0.118	0.045	-0.101	4184
this month	(0.065)	(0.059)	(0.062)	(0.068)+	(0.030)	(0.063)	
Communicated with director	0.900	0.01	-0.001	0.019	0.001	0.009	4184
about students	(0.009)	(0.017)	(0.018)	(0.015)	(0.009)	(0.015)	
Communicated with director	0.991	-0.098	-0.119	-0.143	0.032	-0.131	4184
this month	(0.003)	(0.100)	(0.103)	(0.105)	-0.028	(0.103)	
Average Effect (in sd)		**-0.049**	**-0.106**	**-0.095**	**0.051**	**-0.101**	**4184**
		(0.086)	**(0.084)**	**(0.087)**	**(0.035)**	**(0.084)**	
Panel E: Classroom learning distribution							
Fraction of students scoring	0.580	-0.185	-0.164	-0.131	-0.036	-0.149	4027
beneath the mean last exam	(0.065)	(0.095)+	(0.094)+	(0.105)	(0.025)	(0.096)	

Notes: Col. (1) reports the endline mean in the comparison group.
Standard errors in parentheses are clustered at the CISCO level. Observations are weighted by the probability of selection.
+ significant at 10%; * significant at 5%; ** significant at 1%
All regressions include controls for the month during which the school was surveyed but not shown.
Regressions of teacher practices also have controls for teacher's age, age^2, teacher's salary, and teaching degree.
Treatment 1 is an indicator for whether the school is in a treated CISCO. Treatment 2 is an indicator for whether school is in a treated ZAP. Treatment 3 is an indicator for whether school itself received AGEMAD interventions.

Table 5: School Functioning and Director Practices

Dependent variables	Endline C.Group Mean (1)	Regression (1) Treatment 3 (2)	Regression (1) Treatment 2 (3)	Regression (1) Treatment 1 (4)	Regression (2) Treatment 3 (5)	Regression (2) Any Treatment (6)	Obs. (7)
Panel A: Observations of school at visit							
Students playing in the hall	0.084	0.008	0.017	0.034	-0.018	0.025	1070
	(0.017)	(0.032)	(0.029)	(0.035)	(0.02)	(0.029)	
Teachers in the hall	0.096	-0.009	0.011	-0.003	-0.013	0.004	1070
	(0.018)	(0.029)	(0.028)	(0.033)	(0.02)	(0.028)	
Average Effect (in sd)		**-0.004**	**0.050**	**0.052**	**-0.055**	**0.051**	1070
		(0.087)	**(0.078)**	**(0.104)**	**(0.064)**	**(0.083)**	
Panel B: Observations of classrooms at visit (classroom-level)							
Classroom with a teacher	0.934	-0.02	-0.02	-0.013	-0.004	-0.016	4316
	(0.007)	(0.03)	(0.02)	(0.02)	(0.02)	(0.020)	
Classroom with a lesson in progress	0.843	-0.014	-0.019	-0.01	0.001	-0.015	4316
	(0.011)	(0.06)	(0.04)	(0.05)	(0.03)	(0.046)	
Average Effect (in sd)		**-0.056**	**-0.062**	**-0.038**	**-0.006**	**-0.050**	4316
		(0.097)	**(0.077)**	**(0.082)**	**(0.059)**	**(0.076)**	
Panel C: Director's control and reporting practices							
Teacher absences reported to Cisco	0.265	-0.043	-0.088	-0.076	0.039	-0.082	1202
	(0.026)	(0.125)	(0.130)	(0.128)	(0.04)	(0.129)	
Teacher absences reported to Zap	0.833	-0.068	-0.131	-0.027	0.011	-0.08	1202
	(0.022)	(0.069)	(0.087)	(0.067)	(0.023)	(0.076)	
Director checks student roll call at least weekly	0.231	0.014	0.027	-0.012	0.007	0.008	1202
	(0.024)	(0.085)	(0.060)	(0.062)	(0.035)	(0.060)	
Director records teacher absences	0.901	0.085	0.054	0.07	0.023	0.062	1202
	(0.019)	(0.034)*	(0.036)	(0.035)*	(0.012)+	(0.035)+	
Director records monthly student attendance	0.769	-0.002	-0.027	-0.08	0.051	-0.053	1202
	(0.026)	(0.066)	(0.069)	(0.078)	(0.038)	(0.071)	
Average Effect (in sd)		**0.030**	**-0.057**	**-0.030**	**0.074**	**-0.044**	1202
		(0.092)	**(0.097)**	**(0.090)**	**(0.039)+**	**(0.092)**	
Panel D: Meetings to discuss school matters							
Days elapsed since last teacher meeting	25.538	1.702	2.849	4.975	-2.212	3.906	1195
	(2.710)	(3.863)	(4.131)	(4.251)	(4.47)	(3.445)	
Days elapsed since last meeting with community	51.682	-1.79	-1.713	-6.371	2.291	-4.062	1195
	(5.269)	(8.971)	(7.852)	(7.831)	(6.106)	(7.678)	
No parent conference this past month	0.481	0.114	0.131	0.145	-0.024	0.138	1195
	(0.029)	(0.090)	(0.090)	(0.091)	(0.031)	(0.087)	
Average Effect (in sd)		**0.079**	**0.100**	**0.099**	**-0.020**	**0.099**	1195
		(0.077)	**(0.063)**	**(0.075)**	**(0.060)**	**(0.065)**	
Panel E: Visits from authorities							
Days elapsed since Cisco visit	294.352	-105.423	-75.272	-33.38	-51.06	-54.979	541
	(67.803)	(68.447)	(83.113)	(72.301)	(36.284)	(71.137)	
Days elapsed since Zap visit	88.907	-31.715	-35.608	-20.836	-3.633	-28.208	541
	(10.387)	(8.283)**	(10.739)**	(10.813)+	(5.508)	(9.872)**	
Average Effect (in sd)		**-0.318**	**-0.309**	**-0.169**	**-0.080**	**-0.240**	541
		(0.105)**	**(0.126)****	**(0.117)**	**(0.040)***	**(0.110)***	

Notes: Col. (1) reports the endline mean in the comparison group.
Standard errors in parentheses are clustered at the CISCO level.
+ significant at 10%; * significant at 5%; ** significant at 1%
All regressions include controls for the month during which the school was surveyed but not shown.
Treatment 1 is an indicator for whether the school is in a treated CISCO. Treatment 2 is an indicator for whether school is in a treated ZAP. Treatment 3 is an indicator for whether school itself received AGEMAD interventions.

Table 6: How Frequently Does the Zap Head Interact with Schools of Different Categories?

	Dependent variables					
	Number of Meetings in year 1 *(Comparison Group Mean = 1.582)*			Number of Visits in year 1 *(Comparison Group Mean = 3.098)*		
Schools categorized by	Net Exit	Pass Rate	Repetition Rate	Net Exit	Pass Rate	Repetition Rate
	(1)	(2)	(3)	(4)	(5)	(6)
Exceptional	0.183	-0.22	-0.072	-0.553	-1.228	-0.932
	(0.155)	(0.158)	(0.096)	(0.479)	(0.463)*	(0.520)
Difficulty	-0.161	-0.164	0.011	-1.153	-1.844	-0.883
	(0.162)	(0.112)	(0.127)	(0.426)*	(0.695)*	(0.673)
Disappointing	0.188	-0.255	0	-0.196	-1.397	-0.394
	(0.316)	(0.161)	(0.085)	(0.555)	(0.514)*	(0.619)
Treatment Zap	0.259	0.238	0.365	-0.435	-0.821	-0.347
	(0.344)	(0.337)	(0.203)	(0.941)	-1.13	(1.286)
Exceptional*Treatment Zap	-0.251	0.038	-0.32	-0.365	0.02	-0.563
	(0.358)	(0.296)	(0.173)	(0.711)	-0.725	(0.989)
Difficulty*Treatment Zap	-0.172	-0.269	-0.369	-0.211	0.319	-0.299
	(0.360)	(0.360)	(0.221)	(0.626)	-1.047	(1.091)
Disappointing*Treatment Zap						
	-0.247	-0.213	-0.373	-0.398	0.187	-0.434
	(0.502)	(0.435)	(0.196)	(0.662)	-0.899	(1.069)
Constant	0.75	1.066	0.842	2.072	3.033	2.194
	(0.252)*	(0.326)**	(0.263)**	(0.697)*	(0.898)**	(0.834)*
Observations	1018	785	1020	1018	785	1020

Notes: Each column is a regression of meetings or visits by ZAP head. The regressors include different categories of schools ("exceptional, difficulty, disappointing" with omitted category: satisfactory).
Schools can be categorized by net exit, exam pass rate or repetition rate, hence the corresponding columns.
Net exit can be understood as dropout rate. It is the difference between school enrollment at the beginning and at the end of the year, divided by initial enrollment.
Standard errors reported in parenthesis, clustered at the CISCO level
* significant at 5%; ** significant at 1%
All regressions control for ZAP head's years of experience, experience squared, and having an advanced teaching degree.
Treatment Zaps are those in treatment groups 2 and 3. Year 1 refers to the first year of the interventions, 2005-2006.

Table 7: Treatment Effects on Student Attendance

Dependent variables	Endline C.Group Mean	*Regression (1)*			*Regression (2)*		
		Treatment 3	Treatment 2	Treatment 1	Treatment 3	Any Treatment	Obs.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Student attendance (in %)							
Surveyor reported	80.113	-1.156	-1.94	-1.508	0.567	-1.725	945
	(1.293)	(4.583)	(3.645)	(3.484)	(1.922)	(3.524)	
Classroom count	85.333	2.394	1.897	1.662	0.613	1.783	5097
(classroom-level)	(0.492)	(2.331)	(2.067)	(2.410)	(0.777)	(2.195)	
Director's book record	86.589	4.29	1.67	3.12	1.895	2.397	901
	(0.784)	(1.96)*	(2.310)	(2.050)	(0.643)**	(2.155)	

Notes: Col. (1) reports the endline mean in the comparison group.
Standard errors in parentheses are clustered at the CISCO level. Observations are weighted by the probability of selection.
+ significant at 10%; * significant at 5%; ** significant at 1%
All regressions include controls for the month during which the school was surveyed, and index for school infrastructure
Treatment 1 is an indicator for whether the school is in a treated CISCO. Treatment 2 is an indicator for whether school is in a treated ZAP. Treatment 3 is an indicator for whether school itself received AGEMAD interventions.

Table 8: Treatment Effects on Repetition, Dropouts and CEPE Results

	Endline C.Group Mean	*Regression (1)*			*Regression (2)*		
		Treatment 3	Treatment 2	Treatment 1	Treatment 3	Any Treatment	Obs.
Dependent variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: End-year 1 school performance							
CEPE exam pass rate	0.691	0.004	0.039	0.061	-0.046	0.050	1113
	(0.016)	(0.053)	(0.049)	(0.048)	(0.020)*	(0.047)	
Net exit from Grade 5	0.087	-0.076	-0.063	-0.048	-0.02	-0.056	1113
	(0.008)	(0.019)**	(0.020)**	(0.019)*	(0.019)	(0.018)**	
Net exit from all grades	0.061	-0.017	-0.016	-0.024	0.003	-0.02	1113
	(0.007)	(0.015)	(0.012)	(0.013)+	(0.006)	(0.012)+	
CEPE exam attendance rate	0.899	0.05	0.038	0.033	0.014	0.036	1113
	(0.010)	(0.030)+	(0.030)	(0.028)	(0.02)	(0.027)	
Average Effect (in sd)		**0.174**	**0.183**	**0.206**	**-0.021**	**0.195**	1113
(sign of net exit reversed)		**(0.089)+**	**(0.088)***	**(0.090)***	**(0.038)**	**(0.085)***	
Panel B: Grade repetition rates over time							
Year 1 grade repetition	0.220	-0.002	0.012	0	-0.008	0.006	1118
	(0.006)	(0.022)	(0.022)	(0.020)	(0.008)	(0.020)	
Year 2 grade repetition	0.226	-0.051	-0.026	-0.045	-0.016	-0.035	1085
	(0.007)	(0.019)*	(0.019)	(0.020)*	(0.007)*	(0.019)	

Notes: Col. (1) reports the endline mean in the comparison group.
Net exit can be understood as dropout rate. It is the difference between school enrollment at the beginning and at the end of the year, divided by initial enrollment.
Standard errors in parentheses are clustered at the CISCO level. Observations are weighted by the probability of selection.
+ significant at 10%; * significant at 5%; ** significant at 1%
CEPE is the primary cycle exam at the end of Grade 5
Treatment 1 is an indicator for whether the school is in a treated CISCO. Treatment 2 is an indicator for whether school is in a treated ZAP. Treatment 3 is an indicator for whether school itself received AGEMAD interventions.
Year 1 refers to the first year of the interventions, 2005-2006. Year 2 is 2006-2007.

Table 9: Treatment Effects on Endline Student Test Scores

	Regression (1)			*Regression (2)*	
	Treatment 3	Treatment 2	Treatment 1	Treatment 3	Any Treatment
	(1)	(2)	(3)	(4)	(5)
Dependent variables: Normalized endline test scores					
Total Score	0.095	0.012	-0.013	0.095	0
	(0.11)	(0.09)	(0.10)	(0.050)+	(0.10)
French	0.035	-0.015	-0.017	0.051	-0.016
	(0.11)	(0.10)	(0.10)	(0.05)	(0.10)
Malagasy	0.115	0.046	0.018	0.083	0.032
	(0.10)	(0.09)	(0.10)	(0.040)*	(0.09)
Math	0.095	-0.006	-0.041	0.119	-0.024
	(0.10)	(0.10)	(0.10)	(0.047)*	(0.10)

Notes: N= 21126. Test scores are normalized by subtracting the mean and divided by standard deviation of control group's baseline scores. Standard errors in parentheses are clustered at the Cisco level
Observations are weighted by the probability of selection.
+ significant at 10%; * significant at 5%; ** significant at 1%
Treatment 1 is an indicator for whether the school is in a treated CISCO. Treatment 2 is an indicator for whether school is in a treated ZAP. Treatment 3 is an indicator for whether school itself received AGEMAD interventions.

Appendix Figure 1: Sample School Report Card

MENRS

DIRESEB:
CISCO:
ZAP:

SCHOOL REPORT CARD: YEAR 2005

(Based on the information the school submitted in FFA 2004-2005)

SCHOOL:

SECTOR:

SCHOOL CODE:

ACADEMIC RESULTS

Rate of net exit

	Your school	Your CISCO
CP1		
CP2		
CE		
CM1		
CM2		
Average		

Repetition rate

	Your school	Your CISCO
CP1		
CP2		
CE		
CM1		
CM2		
Average		

CEPE results

	Your school	Your CISCO
CM2 enrollment		
Present at CEPE exam		
Passed the CEPE		
CEPE pass rate		
% of CM2 passed CEPE		

STUDENTS AND TEACHERS

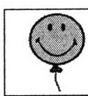
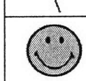
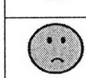
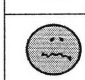
	Your school
Nb. Students	
Nb. Teachers	
Holding classes	
Civil service teachers	
FRAM	
Supplements	
Not holding class	
Total	

	Your school	Your CISCO
Nb. Students/teacher		
% of FRAM (contract) teachers		
% teachers not holding class		

WHERE DOES YOUR SCHOOL STAND IN TERMS OF PERFORMANCE AND RESOURCES?

	Your school relative to others in your CISCO
Rate of net exit	
Repetition rate	
CEPE results	

	Your school	Your CISCO
Rate of net exit Repetition rate CEPE results Nb. Students/teacher		

	Results	Nb. Students/teacher	School Category
	Better than CISCO average	Less than CISCO average	Exceptional
	Better than CISCO average	Better than CISCO average	Satisfactory
	Less than CISCO average	Less than CISCO average	In Difficulty
	Less than CISCO average	Better than CISCO average	Disappointing

Appendix Table 1: AGEMAD Tools for Use by Teachers

<table>
<tr><th rowspan="2">Objective</th><th rowspan="2">Activities</th><th colspan="4">AGEMAD Tool</th></tr>
<tr><th>Code</th><th>Description</th><th>Theme</th><th>Periodicity</th></tr>
<tr><td rowspan="3">Provide overview of the year's curriculum and promote systematic lesson planning</td><td rowspan="3">Plan time use and calendar of lessons
Prepare lessons
Track pupils' reception of the lessons</td><td>E1A/PED</td><td>Bimonthly lesson plan for the entire school year</td><td>Pedagogy</td><td>At start of school year</td></tr>
<tr><td>E1B/PED</td><td>Weekly lesson plan for bimonthly periods</td><td>Pedagogy</td><td>At start of each bimonthly period</td></tr>
<tr><td>E2/PED</td><td>Individual lesson plan</td><td>Pedagogy</td><td>Daily</td></tr>
<tr><td>Put focus on student learning</td><td>Prepare and mark tests
Record test scores
Identify and help lagging pupils</td><td>E3/EVA</td><td>Record of bimonthly test scores
(Cahier de notes de class)</td><td>Evaluation</td><td>After each bimonthly test</td></tr>
<tr><td>Mobilize parental support for academic excellence</td><td>Inform parents about pupils' progress in school</td><td>E4/EVA</td><td>Individual pupil report card
(Bulletin individuel de notes)</td><td>Evaluation</td><td>End of each bimonthly period</td></tr>
<tr><td>Reduce student absenteeism</td><td>Monitor pupils' attendance
Detect possible attendance problems and their causes
Take remedial action</td><td>E5/APP</td><td>Class attendance register</td><td>Time for learning</td><td>Daily</td></tr>
<tr><td>Reduce teacher absenteeism</td><td>Account for teacher absences
Detect possible attendance problems and their causes
Take remedial action</td><td>E6/APP</td><td>Teacher's personal leave/travel record card</td><td>Time for learning</td><td>Each occasion of absence</td></tr>
</table>

Appendix Table 2: AGEMAD Tools for Use by School Directors

Objective	Activities	AGEMAD tool			
		Code	**Description**	**Theme**	**Periodicity**
Ensure proper registration of pupils	Keep up-to-date register of pupils at the school	D1/ADM	*Registre matricule* (National Printer document)	Administration	Each time a new pupil registers
Reduce teacher absenteeism	Monitor presence of teachers	D2/APP	Attendance register	Time for learning	Daily
		D3/APP	Summary table of teacher absences	Time for learning	At end of each month
Reduce pupil absenteeism	Record summary of pupil attendance records Review attendance record and assess possible problems and their causes Take remedial action	D4/APP	Summary table of monthly record of pupil absenteeism	Time for learning	At end of each month
Focus on progress of student learning at the school	Call periodic staff meetings (*Conseil des Maîtres*) Track student learning, note strengths and weaknesses, and plan to implement remedial action as needed	D5/EVA	Summary of student test scores	Evaluation	At end of each bimonthly period
Improve school's internal efficiency	Use school's data on student flow to identify and address possible problems Ensure application of grade to grade transition criteria (i.e., *système de cours*) [a/]	D6/EVA	Student flow table	Evaluation	End of each school year
Strengthen school's-partnership with the local community	Organize meetings with parents and teachers, FAF and FRAM Sensitize parents of pupils at the school Work with parents to prepare a *contrat programme de réussite scolaire* (CPRS)	D7/PART	Community meeting form	Community relations	At each meeting with the community
Improve conditions for learning at the school	Assign teaching duties among staff, and allocate classrooms and teaching materials; and select teachers for in-service training.	D8/PED	Organization of pedagogical arrangements	Pedagogy	At start of school year
Improve performance of the school	Study and display in public area the school report card Update school's performance indicators Discuss the school report card with community and use it to inform development of a school improvement plan (*contract programme*) for implementation	D9/TDB	School report card (*Tableau de bord de l'école*)	Focus on results	At end of school year

a/ Automatic promotion between grades 1 and 2 and between grades 4 and 5.

Appendix Table 3: AGEMAD Tools for Use by ZAP Heads

Objective	Activities	AGEMAD Tool			
		Code	Description	Theme	Periodicity
Improve management of schools in ZAP	Keep up-to-date inventory of schools in the ZAP	Z1/ADM	List of all schools in the ZAP	Administration	At start of school year
Improve the education information system	Ensure timely and complete collection of school data on the census forms	Z2/STA	Checklist of data collection	Statistics	At start and end of school year
Improve pupils' access to pedagogical materials	Determine number of pupils in each functioning school to receive a school kit	Z3/STA	Enrollments by school and distribution of the school kits	Statistics	At start of school year
Strengthen teacher competencies	Take stock of in-service teacher training (*journées pédagogiques*)	Z4/FOR	*Aide mémoire* of in-service teacher training event	Training	After each training event
	Supervise and provide pedagogical support to teachers in schools Provide feedback to school director Track assimilation of training provided Identify potential areas of additional training to help teachers become more effective	Z5/PED	Pedagogical supervision and support form	Pedagogy	After each school visit
Improve management at school finances	Track expenditure against grants provided through the *Caisse Ecole* and *Caisse Cantine*	Z6/ADM	*Caisse Ecole* form	Administration	On each school visit
	Ensure public posting of the expenditures	Z7/ADM	*Caisse Cantine* form	Administration	On each school visit
Strengthen school-community partnership	Maintain relations with the community Visit/meet community partners for education	Z8/PART	Community meeting form	Community relations	On each school visit
Improve performance of schools in the ZAP	Take note of each school's progress from year to year Compare schools in the ZAP, identify lagging schools, and plan extra support and attention for such schools	Z9/TDB	ZAP report card (*Tableau de bord de la ZAP*)	Focus on results	At start of school year

Appendix Table 4: AGEMAD Tools for Use by CISCO Heads

Objective	Activities	AGEMAD Tool			
		Code	Description	Theme	Periodicity
Manage teacher transfers	Take stock of existing deployment of teachers Summarize and process transfer requests	C1/ADM	Teacher transfer master sheet	Administration	End of school year
Improve schooling conditions	Inform the Ministry about requirements for school kits and other school supplies Distribute school kits, textbooks and other supplies	C2/ADM	Pedagogical supplies form	Administration	Before start of school year
Improve management of school finances	Distribute the *Caisse Ecole* and *Caisse Cantine* grants Ensure follow-up on utilization of the grants	C3A/ADM	*Caisse Ecole* follow-up sheet	Administration	End of second bimonthly period
		C3B/ADM	*Caisse Cantine* follow-up sheet	Administration	End of second bimonthly period
		C4A/ADM	*Caisse Ecole* verification form	Administration	On occasion of each visit to ZAP
		C4B/ADM	*Caisse Cantine* verification form	Administration	On occasion of each visit to ZAP
Strengthen teacher competencies	Identify teacher training needs and organize training events Implement post-training follow-up	C5/FOR	Planning sheet for in-service teacher training (*journées pédagogiques*)	Training	Before each training event
		C6/FOR	Follow-up sheet for in-service teacher training	Training	Before and after each training event
	Plan school visits to provide in-service support, giving priority to the schools in difficulty Provide feedback to school personnel based on systematic on-site recording of classroom practices	C7/PED	Class observation grid (*Grille d'observation de classe*)	Pedagogy	On each school visit
Improve performance of schools in CISCO	Evaluate outcomes across ZAPs and schools and identify lagging units for extra support and attention Analyze possible sources of difficulty in lagging units and plan remedial action for implementation	C8/TDB	CISCO report card *(Tableau de bord de la CISCO)*	Focus on results	At start of school year

Chapter 3

Incentives against Corruption in Acute Health Care in Vietnam

3.1 Introduction

Corruption is widely regarded as an important issue in the developing world. The efficiency costs of corruption may be a strong barrier to growth in these countries.[1] Corruption involved in health care provision, in particular, may limit redistribution and weaken service delivery, risking human lives directly. Yet, in the public health sector, illegal informal payments for health care services at public facilities are prevalent in many developing countries (Lewis 2006). These payments are made under the table to doctors and nurses, sometimes to get the most basic services that patients are supposed to receive for free. The poor and credit-constrained who cannot offer such payments may fail to receive timely medical treatments.

This paper studies corruption in situations with high rents to extract, in particular informal payments for acute health care. Let us consider acute conditions like a heart attack, which tend to be observable to both the patient's family and the doctor. Acute patients may be more susceptible to corruption than patients of chronic conditions since they would gain more from a treatment and thus have higher willingness to pay. Meanwhile, acute cases also involve a high mortality risk without timely medical attention, so refusing to treat them can leave serious medical consequences or even cost lives. Bureaucrats might get disutility from letting their patients die, or are sensitive to existing incentives against doing so. It is therefore not clear a priori if acute cases would face more or less informal payments than chronic ones. I demonstrate these possibilities in a simple bargaining model, which predicts that acute patients pay less where the doctor's outside option for

[1] For example, this point is made by Mauro (1995) and Rose-Ackerman (2004).

not treating them is sufficiently low.

To investigate this relationship empirically, I examine micro data on informal payments (or side payments) during inpatient visits in Vietnam. This is a context where the government itself recognizes in its health report that substantial informal payments exist at hospitals and may exceed official payments, though the exact amount is unknown. The Vietnamese government also considers informal payments a sensitive issue that should be further studied to reduce the burden on the poor (the very poor are supposed to be fully subsidized in health care) and to improve service quality (Ministry of Health 2003). The Vietnam National Health Survey 2001-2002 provides a direct measure of corruption at hospitals in the form of illegal side payments to doctors and nurses, above and beyond the official fees set by the government. Exploiting variation across visits within the same hospital, I estimate a large reduction in side payments (9.46 thousand dongs or 18%) associated with acute visits relative to non-acute ones, despite the similar official fees. Acute patients are also 8 percentage points less likely to pay bribes at all. This relationship is strong even at facilities where side payments in non-acute care are large, suggesting that otherwise "greedy" bureaucrats might still refrain from life-risking corrupt activities in this context.

There are several interpretations for the empirical negative relationship between side payment and acute condition. Since acute patients presumably have a higher private value of treatment, the result that they pay a lower price implies sufficiently low expected payoffs to doctors for not treating them. One interpretation of the low payoff is that the doctors themselves get disutility from denying treatment in acute care. For example, altruistic doctors do not like the possibility of letting their patient die. Or doctors who have a private practice are concerned about losing reputation from failures to treat acute cases. Another interpretation of the low payoff is due to external incentives against neglecting acute patients, already enforced in the current institutions. Specifically in Vietnam, promotion reviews for doctors are supposed to take into account successful treatments of acute cases. If well-enforced, this policy gives one reason for doctors not to demand excessive bribes in acute care. The government also has a central audit agency to monitor medical practices at all health facilities. The court system provides another possible channel of incentives for doctors. If the patient's family sues the doctor and the hospital for negligence or malpractice in causing the patient's death or health damages, medical reports and testimonies will be examined in court. Punishment consequences include suspension of the doctor's practice license and monetary compensations. This situation is particularly undesirable since there is no malpractice insurance in Vietnam.

I provide supporting evidence to suggest that response to incentives, among the above possibilities, is one plausible interpretation of the lower side payments observed in acute care. To shed light on this hypothesis, I compare payments by illness condition in presumably high-incentive versus low-incentive facilities. Motivated by the model's prediction, I test for a larger reduction in side payments in acute care at places where we would expect stronger incentives. I first investigate bribe payments in the central cities, Hanoi and HoChiMinh City. This is where the legal system and the media are most accessible, and where the central authorities and auditing bureaus are located. I find that acute patients pay substantially less than non-acute patients in those central cities (by 69000 dongs), compared to the non-central facilities. The differential propensity to pay bribe at all is 20 percentage points. There is no strong reason to believe that doctors in Hanoi and HoChiMinh City have such superior altruism that it drives this finding entirely. It is possible that doctors there have more private practices and care more about reputation, or they are responding to larger disincentives in the central cities for taking bribes from acute patients.

To exploit cleaner variation in incentives, I use the number of supervision visits each facility receives as a measure of monitoring. Among all the commune facilities where this data is available, the interactions of acute cases and audit visits are significantly negative in affecting side payments. That means, the reduction in payments when a case is acute is stronger in well-audited facilities than otherwise. For each additional audit visit, acute patients are 1 percentage point less likely to pay bribe than non-acute patients to the same facility, and pay 0.2 percentage points less bribe as a fraction of total payment. While I am not estimating directly a causal relationship between incentives and bribes, this paper finds that even very corrupt bureaucrats are sensitive to risking people's lives in extracting bribes. Well-monitored bureaucrats seem particularly more sensitive than the others. This result is indicative of doctors' reaction to incentives even in a corruptible environment.

This study contributes to the literature on fighting corruption. While the theory for using monitoring and incentives to control corruption dates back to Becker and Stigler (1974), the set of empirical papers on this topic is limited. Evidence to date suggests that bureaucrats respond to incentives against breaking the rules for money. Di Tella and Schargrodsky (2003) shows that procurement prices reported by hospitals in Buenos-Aires dropped significantly once an auditing system was in place. Olken (2007) provides evidence from a field experiment in Indonesia that top-down monitoring reduces corruption in road projects. Publishing transparent information about funding for education in the newspapers reduced local capture of government funds in Uganda (Reinikka and Svensson 2004).[2]

Aside from that, this paper also relates to the literature on bribes and discrimination in health care. On the question of who is paying bribes, Hunt (2007) documents with data from Peru and Uganda that the rich are more likely to pay bribes in health care, and conditional on that, pay larger amounts.[3] I find similar relationships in my data. On the question of what people are paying for, Thompson and Xavier (2002) argues that bribes are paid in exchange for better quality of care, faster admission in particular. They exploit variation across visits within the same hospital and find that unofficial payments and wait time are negatively correlated in one hospital, and positively correlated in another hospital.[4] This correlation is not strong in either direction in the Vietnam data.

The rest of the paper is structured as follows. The next section presents a simple model of

[2]In terms of cross-country evidence, Yang (2008) finds that countries implementing the private sector monitoring program (called "hiring integrity") tend to collect more import duty later.

[3]Svensson (2003) shows this result in a sample of firms in Uganda.

[4]For a large sample of firms in many countries, Kaufmann and Wei (1999) shows that self-reported time spent with bureaucrats is positively correlated with perceived bribery.

bargaining, highlighting the different payoffs for doctors in acute health care. In section 3.3, I provide background information on the Vietnam health care system and informal payments, and describe the data for analysis. Section 3.4 discusses the estimation strategy and empirical results. The last section concludes.

3.2 A Model of Side Payment and Illness Condition

This simple model describes the Nash bargaining solution to determine prices in a transaction between the doctor and a patient of acute or non-acute condition. The main idea of the model is a tradeoff between the patient's value of treatment and the doctor's outside option in affecting the total surplus from this transaction. The patient has a higher private benefit of being treated for acute than non-acute illnesses. But if the expected payoff to the doctor for not treating acute patients is sufficiently low, the model predicts lower side payments to doctors in acute cases.

Consider bargaining for side payments between a patient of a certain illness and a doctor. There are 2 kinds of illnesses, observable to all parties: acute (A) or non-acute (NA). Both parties have a discount factor β. There are 2 possible time periods for treatment, and by period 2, the doctor has to treat a patient that has been to the hospital. There is no imperfect information or uncertainty, so bargaining is resolved in period 1.

An acute patient has the following private values of the treatment in either period: $b_1^A > 0$ if he receives the treatment in period 1; otherwise if he is treated in period 2, he gets $b_2^A = 0$ since he is likely to die then. For the doctor, he obtains a payment p^A from the patient if he gives the treatment. Otherwise, he is charged with a "fine" of $F > 0$ for not treating an acute patient in period 1. The model's insights remain if instead, the fine happens in period 2 when the patient dies, and F is the expected discounted value of the punishment. This exogeneous fine captures, for example, any disutility of the doctor for letting an acute patient die without offering treatment, loss of profit due to damaged reputation, or a punishment set by the government.

A non-acute patient has the follow private values: b_1^{NA} if he gets the treatment in period 1; otherwise if he is treated in period 2, he gets b_2^{NA}. Let us assume $0 < b_2^{NA} < b_1^{NA}$ since illnesses are better treated early, but non-acute cases are not immediately life-threatening. For the doctor, he receives a payment p^{NA} if he treats the non-acute patient in period 1. Otherwise, his outside option is 0 since in period 2, he has to give the patient the treatment.

Since there is surplus to be shared, consider Nash bargaining as a natural benchmark. Let α denote the patient's bargaining power. The resulting price in an acute case would be the solution to the following problem:

$$\max_{p^A}(b_1^A - p^A - \beta b_2^A)^\alpha (p^A + F)^{1-\alpha} \tag{3.1}$$

$$\max_{p^A}(b_1^A - p^A)^\alpha (p^A + F)^{1-\alpha} \tag{3.2}$$

where the term in the first brackets is the payoff to the patient minus his outside option, and the term in the second brackets is the payoff to the doctor minus his outside option.

Similarly for a non-acute case, p^{NA} solves

$$\max_{p^{NA}}(b_1^{NA} - p^{NA} - \beta b_2^{NA})^\alpha (p^{NA} - 0)^{1-\alpha} \tag{3.3}$$

The Nash bargaining solutions imply that prices are set such that the patient's share of the total surplus is equal to his bargaining power

$$\alpha = \frac{b_1^A - p^A}{b_1^A + F} \tag{3.4}$$

$$\alpha = \frac{b_1^{NA} - p^{NA} - \beta b_2^{NA}}{b_1^{NA} - \beta b_2^{NA}} \tag{3.5}$$

For a given α, we can solve for the prices

$$p^A = b_1^A - \alpha(b_1^A + F) \tag{3.6}$$

$$p^{NA} = (b_1^{NA} - \beta b_2^{NA})(1-\alpha) \tag{3.7}$$

The condition with which $p^{NA} > p^A$ is

$$(b_1^{NA} - \beta b_2^{NA})(1-\alpha) > b_1^A(1-\alpha) - \alpha F \tag{3.8}$$

$$F > [b_1^A - (b_1^{NA} - \beta b_2^{NA})]\frac{(1-\alpha)}{\alpha} \tag{3.9}$$

The left hand side is the cost to the doctor of not treating acute cases. The first term on the right hand side denotes the difference in values to the patient of getting treated now (rather than his outside option) between the acute and non-acute cases. The second term is the relative bargaining weights. When F is relatively large and this inequality is satisfied, acute patients have to pay less than non-acute patients.

In the framework thus far, bargaining is resolved and treatment always takes place in period 1. In reality, some patients have to wait for treatment, possibly due to constrained capacity at the hospital. It is worth discussing briefly what implications this has for the model. Suppose each doctor has one acute patient and one non-acute patient, and can only treat one person in each period. If he treats the non-acute first, his payoff is $p^{NA} - F$. If he treats the acute first, his payoff is p^A since in period 2, the non-acute patient has to be treated and does not need to pay. Given the expression for prices in equations 3.6 and 3.7, the doctor is better off taking his acute patient first. The main insight is that the non-acute patient, treated later, now pays a lower price than what he would have to pay if treated in period 1. Thus, when we later look at the difference $p^{NA} - p^A$ in the data where capacity constraints play a role, this empirical estimate will be less than the difference $p^{NA} - p^A$ in an environment without such constraints.

To sum up, while acute cases generally have a higher private benefit of treatment, they may still face lower side payments in this model as long as there is sufficient cost to the doctor of letting these patients die. The model also predicts that the difference in side payments $p^{NA} - p^{A}$ is larger where F is higher. In the remaining sections, I will turn to the data from Vietnam to examine the relationship between bribe payment and disease condition.

3.3 Background and Data

3.3.1 Health Care in Vietnam

The health care system in Vietnam has undergone major reforms since independence. Up until 1989, health care was fully subsidized by the socialist government. In the midst of the liberalization process, the government legalized private practice and started allowing partial fee collection by hospitals to cover operational costs. Decree 95 in 1994 permits public hospitals to collect fees for their own profit, under the condition that this fee schedule is within the official range of user fees and approved by the Ministry of Health.[5]

Today, health care providers include public and private hospitals, traditional healers, and pharmacists. This paper focuses on public hospitals since this is where illegal side payments, above and beyond the permitted official fees, are most common. The public system is hierarchical, with the Ministry of Health in charge of making policies and managing service delivery. At the grassroots level, a commune health care team typically consists of a doctor (or more if larger communes), a nurse and/or a midwife, and village health workers. This network covers all communes and wards throughout the country. Relative to other developing countries in the region, Vietnam has an extensive commune-based health system and delivers many services at the front line. At higher levels, there are usually one hospital per district, one or more provincial hospitals, and some spe-

[5] See Vietnam Health Report 2002 for more information on the health care system and health status in Vietnam (Ministry of Health 2003).

cialized health facilities. Overall, Vietnam has approximately 25 hospital beds and 6 physicians per thousand population (World Health Organization 2008).

Base salaries of doctors and nurses at public hospitals are set by the government. The official salary for a physician is $30-50 per month, and higher for management positions (real GDP per capita was $400 in 2001). In addition, they received premiums for night shifts and for working in remote areas or toxic environments. These premiums are again determined by the government.

In terms of financing, total health spending in the country is roughly 5-6% of GDP. Health expenses were at $27 per year per person in 2005. Health expenditures come from the government budget (26%), out-of-pocket payments (64%), and the rest from insurance plans (World Health Organization 2008). Health insurance was first established in 1992. Nowadays, insurance is mandatory for government employees and deducted directly from their paychecks. Others may purchase voluntary health insurance. The very poor, disadvantaged population, and war veterans receive state subsidies for health care either in the form of free insurance cards or reimbursement for hospital expenses. The health survey data that I exploit in this paper, described below, indicate that one third of the patients have some type of health insurance.

3.3.2 Informal Payment

While the Ministry of Health sets the official fees for different kinds of illnesses, anecdotal evidence suggests that patients pay extra to jump the queue or get the medical staff's attention. The law (Instruction 08/BYT-CT from the Ministry of Health, Jan. 2004) prohibits hospitals from charging any fees other than the official fees as approved by the government. If caught, punishments to doctors range from warning to dismissal from the job. Still, we see deviations from this regulation in the data, thus suggesting there might be corruption and bribe taking. Informal payments may be a significant contribution to the revenues of hospitals and health providers. To get a sense of the magnitude, households report paying 14 times more at public facilities than what the government reports receiving as fees. Taking into account unofficial payments, out-of-pocket financing might actually account for 80% of total health expenditure (Adams 2005).

3.3.3 Data

I use the Vietnam National Health Survey 2001-2002, conducted by the General Statistics Office to collect nationally representative data on health status and health care usage. The entire survey covers 36000 households in 406 urban and 794 rural communes. The inpatient module, relevant for this paper, is part of the household questionnaire that the surveyor used to interview household members at home. It asks for information on all inpatient visits by any member of the household during the 12-month period prior to the interview.[6] If the patient undergoing inpatient care was more than 16 years old and available, he or she was interviewed directly; otherwise, the household head or the next-best respondent answered the survey on their behalf. 7438 households report hospital use for inpatient care, and on average, 1.3 visits per household. Respondents report a series of answers in connection with each visit: type and level of facilities, length of hospital stay, wait time before admission, and official fees and side payments to doctors and nurses. Data on the broad type of illness (acute, injury, or chronic) was collected, without any further details on each illness condition.

The key variable is side payment, which records households' response to the following interview question: "How much did you pay informally to the employees of this facility for this inpatient visit?" This measure of unofficial payments includes the monetary value of in-kind gifts (in Vietnam dongs). If the patient's family did not pay any side payments, the variable was recorded as zero. I observe who does or does not exchange (positive) side payments, and, conditional on paying, the amount in the transaction. I also report the logarithm of side payment amount since its distribution is skewed. To minimize underreporting, the respondents had been told that this health survey was confidential and that their identity would be detached from the answers. If there was systematic underreporting across the board, what this paper is interested in, the difference between side

[6]This survey has a module on outpatient care, but side payments in that setting are rare and of small amounts.

payments for acute and non-acute care, is less subject to this problem than payment levels. Since some respondents might underreport everything as zero, the estimated difference that I observe in the data might be smaller than the latent difference. The survey also asks if some side payment takes place before, during, or after treatment, but only a quarter of the respondents answer this question. For this sub-sample, side payments seem to take place at all three stages, slightly more frequently after the treatment, but I do not know precisely how much is exchanged at each stage.

My analysis focuses on public facilities, which handle 96% of inpatient visits. In order to identify each health facility, I categorize visits into facility identifiers by grouping all visits to the commune (district, province) facility of people in the same commune (district, province). Facility identifiers allow me to include facility fixed effects in my regressions to look at variation across visits within the same hospital. I define a variable called "central cities" as the indicator for the transactions taking place in Hanoi and HoChiMinh City. These two are principal cities with large population sizes, administratively equivalent to a province in Vietnam. The central government's operations and judicial institutions are also housed here.

Another variable of interest to this study is the extent of top-down auditing that would presumably affect bribe payments in equilibrium. The Vietnam National Health Survey collected data on grassroots public health care via a commune health facility questionnaire. In this module, I observe the number of supervision and inspection visits that each facility has received over the 12 months prior to the interview, i.e. the same time period as the data on payments for inpatient care. Typically during these visits, auditors from the Ministry of Health or the provincial Health Bureau may inspect the facility's finances, medical reports, general management, or infrastructure. Places range from one to 80 visits received over one year, with 1 visit per month on average. This data is available for facilities at the commune only.

Table 1 displays a few descriptive statistics on the hospital visits in my sample. On average, inpatient-care clients wait for 40 minutes before admission for a clinical intervention, but about 40% report they do not face any wait time. Patients stay in the hospital for 8 days on average. Half of the visits are for chronic cases such as osteoporosis, hypertension, and tooth decay. Acute diseases such as obstetric complications or heart stroke account for 40% of the visits. Roughly 22% of these visits involve a positive side payment to doctors and nurses. The mean side payment amount is 24000 dongs while the mean official payment to the facility is 367000 dongs. In particular, Figure 1 shows the mean side payment in the whole sample by illness category. Acute cases are less likely to pay and pay less than non-acute cases on average. I will investigate this relationship in more detail in the following section.

The mean annual expenditure per capita for the patient's household is 3652000 dongs, i.e. $281. Household expenditure is constructed using information on a wide range of food expenditure items and asset items collected at the household during the same interview. Per capita expenditure is then the household's total expenditure divided by the household size. Log expenditure will be used in the analysis as a proxy for permanent income.

3.4 Estimation Strategy and Results

This paper studies side payments for different illness categories to see, in particular, if acute patients with high willingness to pay actually have to pay more for treatment than non-acute ones. I run the following regression to estimate this relationship:

$$p_{ihf} = \alpha + \gamma * Logexp_h + \beta_1 * Acute_{ihf} + \beta_2 * Injury_{ihf} + \delta X_{ihf} + \eta_f + \varepsilon_{ihf} \qquad (3.10)$$

where p_{ihf} is an outcome variable for visit i of household h to facility f. I use four measures of side payment as dependent variables: (i) side payment amount as reported in the interview, (ii) an indicator for paying any side payment at all (any strictly positive amount reported), (iii) logarithm of side payment, and (iv) side payment as a share of total payment. The fourth measure, side payment divided by the total amount paid to the facility, gives us a sense of bribes as a markup over the official price. The average share is 11%. Regressions of log side payment have a smaller number of observations as they are restricted to the sample with positive bribes. These regressions should be interpreted with caution due to the sample selection. $Acute_{ihf}$ and $Injury_{ihf}$ are dummies for the corresponding illness category associated with each visit. The omitted category is chronic cases. The coefficients of interest are the β's. Differences in family wealth might drive differences in payment across illness types if, for example, the rich are less likely to catch acute diseases and more likely to pay side payments. I control for this possibility by including log expenditure of the household $Logexp_h$ as a covariate in all the regressions. X_{ihf} refers to other control variables. I include official payment in all the regressions of side payment, except when the dependent variable is side payment as a share of total payment. I will also report some specifications that control for the length of hospital stay since it may be a proxy for severity. However, bribe payments may affect the quality of care and, in turn, how long the patient has to be hospitalized, so this specification is to be interpreted with caution. To account for omitted facility characteristics such as medical quality or location, my preferred specification of equation 3.10 includes facility fixed effects η_f. The estimation of β's comes from variation across visits within health facility. I report robust standard errors in all the regressions.

Aside from side payment, it is useful to check on allocative efficiency. Constrained resources are common at Vietnamese hospitals. Patients and families often huddle in the wait room and outside before and during the treatment. From a social welfare's point of view, acute and injury cases presumably have the highest social benefits of medical treatment and should be the first ones to receive treatment. My approach is to test if acute and injury cases actually have to wait less in equilibrium. I estimate equation 3.10 for a different dependent variable: time spent waiting in the admission department. Part of this wait time is natural queueing time for a certain illness, and part of it may represent unnecessary "red tape." As this variable is rightly skewed, I also report a log transformation of wait time.

3.4.1 Acute Conditions

Despite having higher perceived private benefits of treatment, acute cases tend to pay smaller bribes than non-acute ones in my data. Table 2 reports the estimates from running equation 3.10. Each column is a regression, with the main dependent variables being the four measures of side payment. For each dependent variable, I present the results from an OLS estimation without facility fixed effects in the first column, then with the fixed effects. In these regressions, most of the OLS results are not very different from the fixed effects. The coefficients on "Acute" are consistently negative and statistically different from zero. Non-acute and non-injury patients pay 28.13 thousand dongs on average as side payments to doctors and nurses. Relative to those chronic cases brought to the same facility, acute visits pay on average 9.46 thousand dongs less in bribe (column 2). In particular, they are 8 percentage points less likely to pay bribe, or if they do, pay 18% less. As displayed in column 11, side payment in acute situations is 2 percentage points lower as a share of total payment. These findings are robust to controlling for the length of hospital stay. Injury cases are also 6 percentage points less likely to pay, and pay 6 percentage points less as a share of total payment. However, the payment amount by injury patients is not robustly and significantly smaller than chronic cases.

The results on wait time are indicative of hospitals' sorting behaviors. As shown in column 2 of Table 3, acute cases wait 18 minutes less than chronic cases in the same facility, injury less by 32 minutes. The logarithm specification in columns 4-6 gives similar implications of the results. This is what we expect if hospitals triage patients in the queue, and allocate urgent cases to treatment first.

The coefficients on log expenditure in Tables 2 and 3 imply that rich patients tend to pay more side payments and receive medical interventions sooner. The elasticity of side payment with respect to household expenditure is 0.27 (column 8 of Table 2). The elasticity of wait time with respect to household expenditure is -0.13 (column 5 of Table 3). This relationship is robust throughout the different specifications presented in this paper. An interesting observation in this data is that the tendency for the rich to pay more and wait less is strong among chronic illnesses; it is much weaker or non-existent among acute or injury cases. One possible interpretation is that neglecting acute patients would cost the doctor so much that any discrimination, if ever existing in regular practice, is minimized in acute health care.

3.4.2 High-Bribe Locations

This negative relationship between side payment and acute condition in the Vietnam data still holds at very "greedy" facilities, and is not driven by what happens at places that do not usually collect much bribe anyway. One way of identifying presumably highly corrupt places is to look at the regular bribes collected at each location. I define high-bribe locations, denoted by an indicator $Highsp$, as facilities where side payments in non-acute, non-injury cases exceed the median positive amount out of all facilities. The number of observations is slightly smaller for this variable since some facilities do not have non-acute, non-injury cases reported over this time period. I estimate the following equation with illness types interacted with high-bribe locations, again with facility fixed effects.

$$p_{ihf} = \alpha + \gamma Logexp_h + \beta_1 Acute_{ihf} + \beta_2 Injury_{ihf} \tag{3.11}$$

$$+\theta_1 Acute_{ihf} * Highsp_f + \theta_2 Injury_{ihf} * Highsp_f + \delta X_{ihf} + \eta_f + \varepsilon_{ihf} \tag{3.12}$$

The coefficient of interest θ_1 tells us how the difference in payment for acute and non-acute cases varies at high-bribe versus low-bribe facilities.

Table 4 reports the estimations of equation 3.11, all with facility fixed effects. Again, richer patients are associated with higher side payments and lower wait time. In terms of side payment, the coefficients on the individual term "Acute" are mostly close to zero and statistically insignificant. This indicates that in low-bribe locations, acute patients tend to pay the same amount of bribe as non-acute cases on average. But the coefficients on acute interacted with high-bribe locations are largely negative and significant at the 1% level. In high-bribe locations relative to low-bribe locations, acute patients pay much less than non-acute patients, by 31000 dongs on average. They are 15 percentage points less likely to pay bribes at all, and bribes as a share of the total payment is 5 percentage points smaller. The log side payment regressions in columns 9-10 have relatively large standard errors due to the small sample size. These results are not driven by what happens at commune-level versus district or provincial hospitals since they are robust to controlling for illness conditions interacted with facility type. Acute patients pay less than non-acute ones even at otherwise high-bribe hospitals. Thus, greed in terms of bribes collected in one type of care does not necessarily mean greed in the other.

For injury cases, the evidence is pointing to the same direction. In low-bribe facilities, injury patients are less likely to pay bribes, but tend to pay roughly the same amount as chronic cases on average. In high-bribe facilities relative to low-bribe ones, they are 11 percentage points less likely to pay bribes, and bribes as a share of the total payment are 7 percentage points smaller. Both of these estimates are statistically significant.

In terms of wait time before admission, both acute and injury visits to high-bribe facilities are still correlated with earlier treatment than chronic visits. Columns 1-4 of Table 4 report the results from running equation 3.11 with wait time as the dependent variable. At high-bribe locations, acute cases wait -12.203 - 14.15 = -26.353 minutes less, injury -30.904 - 5.79 = -36.694 minutes less, than chronic ones.

3.4.3 Interpretation

These findings that acute patients pay less bribes than non-acute counterparts, even at usually high-bribe facilities, are intriguing in several ways. One possible interpretation of side payments is that they represent mostly gift giving from the patient's side. The Ministry of Health claims that many patients offer informal payments as gifts to health care staff, and this gesture has its cultural respect that is inherent to Vietnamese people (Ministry of Health 2003). However, the results here do not support this claim. If side payment is mainly a manner of gratitude and does not affect the doctor's efforts, we should expect more (or at least equal) gifts from acute patients than others since they would be particularly grateful for having their lives saved. I find the opposite, suggesting that side payment in this context might be a price for treatment.

Considering side payment as a price in the transaction, there would generally be more surplus to extract from acute or injury cases than otherwise, if the doctor has the same outside option. This is because acute and injury patients tend to have high private values of treatment. But I find instead that acute patients pay less bribes than chronic patients. Since the official payments for acute visits are slightly less than chronic cases within the same hospital, total payments for acute care are 28% less. These results suggest that the doctor's payoff if not engaged in treating acute patients is sufficiently low. One possibility is that doctors, even those accepting high side payments in non-acute care, dislike risking acute patients' lives. They might not mind taking bribes from patients with a minor toothache, but may be altruistic toward those with a heart attack. As in the model's framework, the higher is the disutility to the doctor, the lower is the price charged for acute patients. Alternatively, some doctors with a private practice might worry about damaged reputation if they refuse to treat acute patients. Outsiders with imperfect information may attribute the acute patient's death to the doctor's lack of ability.

Another interpretation of the lower bribes observed in acute care is that doctors are responding to existing incentives against neglecting acute patients. In Vietnam, the prevalence of illegal side payments suggest that incentives against taking bribes in general health care are weak. However, incentives against taking bribes in life-threatening situations may be much stronger. Doctors' promotion decisions,[7] in theory, are made based on qualifications and performance, such as successful treatments of difficult cases. Therefore, failure to treat acute cases would cost them more

[7] In the Vietnamese context, this includes promotion to higher positions within the current hospital, transfer to another preferred facility, or scholarships for further training abroad in a developed country, all of which would result in higher expected life-time earnings.

than failure to treat non-acute ones. The court system provides another channel to hold doctors accountable. If the health provider insists on bribes in order to treat the patient, an acute patient who does not pay has a risk of dying. If the patient's family files a law suit for irresponsibility or malpractice, the doctor and the hospital will have their case examined in court, and often covered in the news where the judicial system and the media are accessible.[8] Even when the hospital and the patient's family settle outside of court, the health providers' reputation is still affected. In addition, the Ministry of Health has an auditing division to directly monitor medical practices at health facilities across the country. I will explore further in the next subsection the possibility that incentives matter and show some supporting evidence that this interpretation is plausible.

It is not straightforward to make the same argument for injury cases. Official payments for injury visits are actually higher than chronic cases within the same hospital, so total payments for injury care are higher by 24%. That is, injury patients pay more in total than chronic patients to the same health facility, which might already reflect their higher willingness to pay.

3.4.4 High-Incentive Locations

This subsection provides some suggestive evidence that lower side payments for acute care may be a sign of the doctor's responses to incentives against neglecting these at-risk patients. I compare payments by illness condition in presumably high-incentive versus low-incentive locations. The model predicts the difference in payment to be larger at facilities where the incentives are expected to be stronger. I investigate this hypothesis by estimating an equation similar to equation 3.11, but the interaction terms are between illness categories and high-incentive locations. I expect the

[8] Many news articles cover cases of doctors and hospitals being sued, mostly in the two central cities–Hanoi and HoChiMinh City. The exact statistics on these law suits are not available, so I cite a few examples here. Mrs. Vo Thi Yen Phi sued HoChiMinh City Medical School Hospital in 2007, after her husband's death post a nose surgery; the case was brought to a district court since negotiations with the hospital failed. Hoan My hospital, HoChiMinh City, had to suspend all of its stent procedures in 2006 as a result of an investigation by the Science Committee of HoChiMinh City, during a law suit by the family of a patient killed after a stent procedure. The Civil Court of Hanoi settled on a negligence case against Viet Phap Hospital in Hanoi in 2003 after the death of new-born twins from obstetric complications, whereby the hospital had to provide monetary compensations (VnExpress 2008).

coefficients on acute visits interacted with high-incentive locations to be negative.

I will present the results using two proxies for high-incentive locations. The first one is an indicator for the central cities, Hanoi and HoChiMinh City. Recall that these cities are close to the government's central offices and the court system. It is risky to neglect acute patients here as they are likely to pursue legal actions. Hospitals in these central cities are also under close scrutiny by the Ministry of Health. Doctors presumably face high potential punishments, delayed promotion, or loss of reputation if denying acute care. Second, I use the audit frequency at each facility as a measure of the amount of monitoring this facility receives. Facilities with more audit visits are associated with lower side payments overall.

I find that the reduction in side payment associated with acute cases is larger in the central cities' facilities than in non-central ones. Table 5 reports how the differential side payments for acute illnesses vary from central to non-central locations. For each dependent variable, I present the results from two estimations, with and without interactions with high-bribe locations and interactions with an indicator for provincial/district facilities. The coefficients on "Acute" imply that in non-central cities, acute patients are 7 percentage points less likely to pay bribes, and pay slightly less on average. In the central cities, visits for acute care pay even much less than those for chronic care, by -5.84 - 69.42 = 75.26 thousand dongs. Acute patients at the central cities' facilities are 27 percentage points less likely to pay bribes than non-acute patients, and pay 9 percentage points less side payment as a share of total payment. As expected, the coefficients on acute interacted with central locations are negative and statistically significant, except for the log side payment regressions with a smaller sample in columns 9-10. This finding holds true whether or not the regressions include additional interaction terms. Similar to the evidence in Table 4, the coefficients on acute interacted with high-bribe locations remain negative and statistically different from zero in columns 6, 8, and 12 of Table 5. I cannot make the same case for injury patients, however. The difference in side payments for injury conditions versus chronic ones does not change from non-central to central locations.

As presented in columns 1-4 of Table 5, acute cases wait 18 minutes less at non-central facilities, but this effect seems to disappear at central facilities. One possible explanation is that large hospitals in the central cities face excessively high demand. Their triaging system in place is not very efficient to sort urgent cases to faster care.

Aside from stronger expected incentives, the central cities might also have a larger market for private practice, affecting doctors' payoffs if acute patients' death damages their reputation. To get a better picture of incentives, I use a direct measure of monitoring. Table 6 reports the results on facilities' audit frequency, i.e. the number of supervision visits received over a year. Note that this measure is available only for the subsample of commune facilities, thus the smaller number of observations in these regressions. The first interesting observation is that the coefficients on "Acute" are small and not always distinguishable from zero. Acute patients pay about the same as non-acute ones at facilities with the lowest supervision. However, in highly audited facilities, we observe a significantly larger reduction in acute patients' tendency to pay side payment and in how much they pay informally as a share of total payment. For each additional audit visit a facility receives, acute patients are 1 percentage point less likely to pay bribe than chronic patients to that same facility, and pay 0.2 percentage points less bribe as a fraction of total payment. Both of these estimates are significant at the 1% level. Places with one more supervision visit per year are associated with 185 dongs less side payment on average in acute relative to non-acute cases. This estimate in column 5 is not precise (p-value = 0.116). While the findings here do not identify a causal relationship, it is suggestive of incentives for bureaucrats at work. Acute patients pay less where doctors face stronger monitoring, thus higher expected punishment for refusing acute care. As shown in columns 6, 8, and 11, the negative coefficients on acute interacted with high-bribe locations remain significant in this subsample. The results on injury incidences are again weak; none of the coefficients on injury interacted with audit visits are statistically significant. The difference in side payments for injury conditions versus chronic ones does not change across facilities of different audit frequencies.

3.5 Conclusion

Corruption in general health care can constrain the extent of the government's service delivery and redistribution. In principle, primary health care consultations in Vietnam are free for the poor; they are either exempt or provided with free insurance cards. But in reality, this purpose is partially defeated if doctors and nurses can request unofficial side payments before treatment. Acute care, however, is particularly urgent and life-threatening. Even if the public health environment is overall corrupt, bureaucrats might still charge less bribes from acute patients than from chronic patients, despite the conjecture that these cases entail high rents that bureaucrats can extract.

In this paper, I study informal payments to doctors and nurses for acute care in Vietnam. Acute care is of interest since the patients tend to have a high private value of treatment. I first develop a simple model with the insight that the doctor's expected payoff if neglecting acute cases may be so poor that they demand less bribes than in non-acute cases. Using data on inpatient care usage in Vietnam, I exploit variation across visits within hospital and find that acute patients are 8 percentage points less likely to pay anything above and beyond the official fees to doctors and nurses. If they do, they pay 18% less side payments than non-acute patients. These findings are strong even in "greedy" places with relatively high bribes collected among patients of chronic conditions.

While there can be several interpretations for this negative relationship between bribes and acute condition, this study provides supporting evidence that doctors may be sensitive to incentives against neglecting acute patients. I find that facilities that receive more monitoring in terms of audit visits are associated with a larger reduction in bribes paid by acute visitors. The doctors seem to respond more where the incentives are expected to be higher. At presumably high-incentive facilities such as those in the central cities, the differential bribe payment between acute and non-acute cases is also larger than that at non-central locations.

In conclusion, the data from Vietnam suggests that corrupt activities that risk lives, such as those in acute health care, are limited even in a potentially corrupt institution. One plausible reason explored in this paper is that bureaucrats, even very greedy ones, may be responding to incentives against such activities.

References

ADAMS, S. (2005): "Vietnam's Health Care System: A Macroeconomic Perspective," Discussion paper at the International Symposium on Health Care in Asia, Hitotsubashi University, Tokyo.

BECKER, G., AND G. STIGLER (1974): "Law Enforcement, Malfeasance, and Compensation of Enforcers," *Journal of Legal Studies*, 3(1), 1–18.

DI TELLA, R., AND E. SCHARGRODSKY (2003): "The Role of Wages and Auditing during a Crackdown on Corruption in the City of Buenos Aires," *Journal of Law and Economics*, 1, 269–292.

HUNT, J. (2007): "Bribery in Health Care in Peru and Uganda," NBER Working Paper 13034.

KAUFMANN, D., AND S.-J. WEI (1999): "Does "Grease Money" Speed up the Wheels of Commerce?," NBER Working Paper 7093.

LEWIS, M. (2006): "Governance and Corruption in Public Health Care Systems," Center for Global Development Working Paper 78.

MAURO, P. (1995): "Corruption and Growth," *Quarterly Journal of Economics*, 110(3), 681–712.

MINISTRY OF HEALTH (2003): *Vietnam Health Report 2002.* Medical Publishing House.

OLKEN, B. (2007): "Monitoring Corruption: Evidence from a Field Experiment in Indonesia," *Journal of Political Economy*, 115 (2), pp. 200–249.

REINIKKA, R., AND J. SVENSSON (2004): "The Power of Information: Evidence from a Newspaper Campaign to Reduce Capture," World Bank working paper 3239.

ROSE-ACKERMAN, S. (2004): "The Challenge of Poor Governance and Corruption," mimeo, Copenhagen Consensus.

SVENSSON, J. (2003): "Who Must Pay Bribes and How Much? Evidence from a Cross Section of Firms," *Quarterly Journal of Economics*, 118, 207–230.

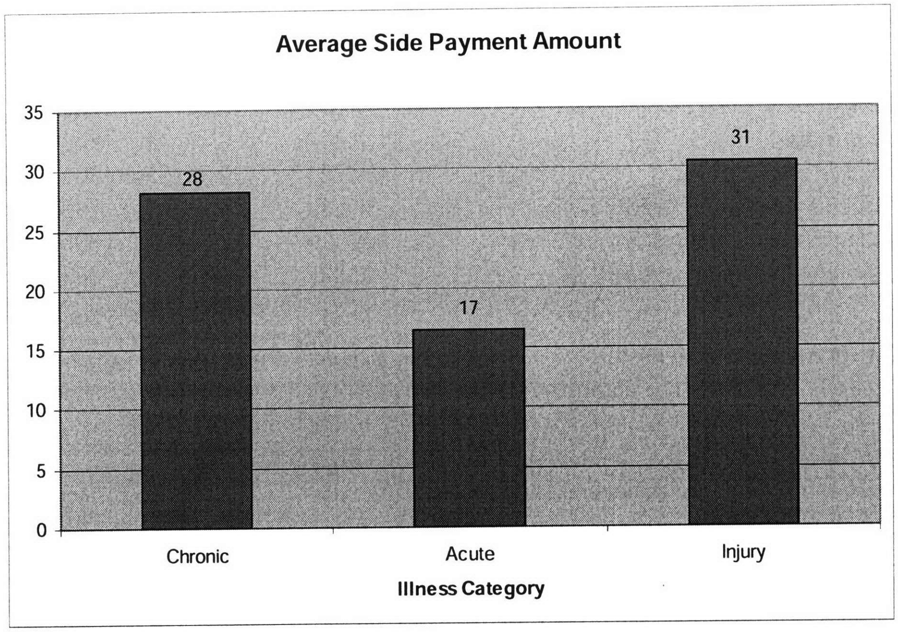
THOMPSON, R., AND A. XAVIER (2002): "Unofficial Payments for Acute State Hospital Care in Kazakhstan. A Model of Physician Behavriour with Price Discrimination and Vertical Service Differentiation," LICOS Working Paper 124.

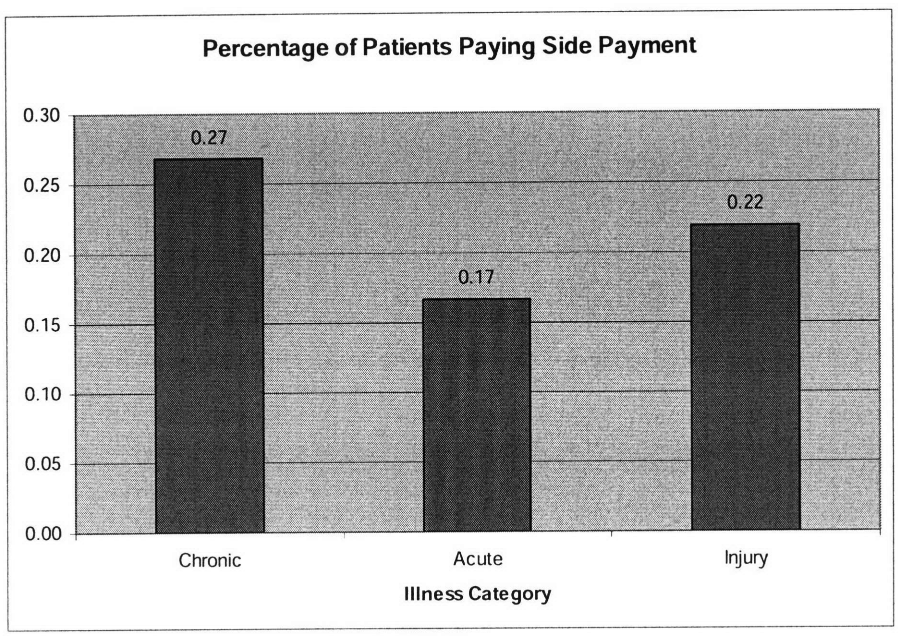
VNEXPRESS (2008): "Vietnam News Database," http://vnexpress.net/GL/Phap-luat/.

WORLD HEALTH ORGANIZATION (2008): "WHO Statistical Information System: Core Health Indicators," Discussion paper.

YANG, D. (2008): "Integrity for Hire: An Analysis of a Widespread Customs Reform," *Journal of Law and Economics*, Vol. 51, No. 1.

Figure 1: Side Payment for Different Illness Categories





Notes: Side payment amount is measured in 1000 dongs.
13000 dongs ~ 1 USD

Table 1: Descriptive Statistics

	Mean	Std. Dev.	Obs
	(1)	(2)	(3)
Wait time (minutes)	39.32	239.28	8085
Length of hospital stay (days)	7.83	8.88	8088
Acute case	0.40	0.49	8096
Injury case	0.12	0.32	8096
Official payment to facility	366.89	1721.10	7850
Side payment amount	23.73	121.24	8060
Pay (positive) side payment	0.22	0.42	8096
District and provincial hospitals[a]	0.84	0.37	8096
Facilities in central cities[b]	0.05	0.22	8096
Household annual expenditure per capita	3652.07	3614.61	8096

Notes: All payment variables are in 1000 dongs. 13000 dongs ~ 1 USD

(a) Omitted category: Commune facilities (public facilities only)

(b) Central cities are Hanoi and HoChiMinh City

Table 2: Side Payments for Different Illness Categories

Dependent variables	Side Payment Amount (*Chronic Cases Mean = 28.13*)			Pay Side Payment (*Chronic Cases Mean = 0.27*)			Log Side Payment (*Chronic Cases Mean = 4.05*)			Side Payment as % of Total (*Chronic Cases Mean = 0.12*)		
	OLS	Fixed Effects	Fixed Effects	OLS	Fixed Effects	Fixed Effects	OLS	Fixed Effects	Fixed Effects	OLS	Fixed Effects	Fixed Effects
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Log expenditure	27.97	25.274	26.752	0.093	0.064	0.066	0.302	0.266	0.279	0.036	0.038	0.042
	(6.668)**	(10.363)*	(10.916)*	(0.010)**	(0.012)**	(0.012)**	(0.058)**	(0.085)**	(0.085)**	(0.006)**	(0.008)**	(0.008)**
Acute	-8.969	-9.46	-7.222	-0.096	-0.077	-0.075	-0.034	-0.179	-0.187	-0.021	-0.018	-0.014
	(3.790)*	(3.689)*	(4.4)+	(0.010)**	(0.011)**	(0.011)**	(0.06)	(0.077)*	(0.078)*	(0.007)**	(0.008)*	(0.008)+
Injury	0.224	-1.002	1.165	-0.05	-0.057	-0.055	0.02	-0.068	-0.066	-0.046	-0.062	-0.059
	(4.60)	(5.76)	(5.44)	(0.016)**	(0.016)**	(0.016)**	(0.08)	(0.10)	(0.10)	(0.008)**	(0.009)**	(0.009)**
Official payment	0.008	0.008	0.006	0	0	0						
	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)						
Log official payment							0.306	0.298	0.282			
							(0.022)**	(0.034)**	(0.034)**			
Length of hospital stay			1.561			0.002			0.007			0.004
			(0.749)*			(0.001)*			(0.00)			(0.001)**
Constant	-200.348	-178.132	-202.872	-0.488	-0.259	-0.287	-0.11	0.28	0.214	-0.172	-0.185	-0.246
	(54.575)**	(84.610)*	(94.409)*	(0.077)**	(0.100)**	(0.101)**	(0.47)	(0.73)	(0.73)	(0.049)**	(0.065)**	(0.065)**
Observations	7826	7826	7820	7850	7850	7844	1424	1424	1423	6073	6073	6070

Notes: Omitted illness category is chronic cases, for which the mean side payments are reported at the top of Table 2.
All payment variables are in 1000 dongs. 13000 dongs ~ 1 USD
"Pay side payment" is an indicator for paying a positive amount
Fixed effects regressions include facility fixed effects
Robust standard errors in parentheses
+ significant at 10%; * significant at 5%; ** significant at 1%

Table 3: Wait Time before Admission for Different Illness Categories

Dependent variables	Wait (in minutes) (*Chronic Cases Mean = 49.63*)			LogWait (*Chronic Cases Mean = 3.2*)		
	OLS	Fixed Effects	Fixed Effects	OLS	Fixed Effects	Fixed Effects
	(1)	(2)	(3)	(4)	(5)	(6)
Log expenditure	0.94	-12.901	-11.791	0.052	-0.131	-0.124
	(5.43)	(7.93)	(7.91)	(0.04)	(0.053)*	(0.053)*
Acute	-16.60	-18.08	-16.562	-0.137	-0.101	-0.082
	(6.089)**	(6.952)**	(6.831)*	(0.041)**	(0.044)*	(0.044)+
Injury	-20.82	-32.363	-31.28	-0.223	-0.312	-0.296
	(8.048)**	(10.253)**	(10.188)**	(0.070)**	(0.073)**	(0.073)**
Length of hospital stay			1.131			0.01
			(0.387)**			(0.002)**
Constant	40.94	154.28	135.813	2.778	4.239	4.094
	(43.38)	(64.641)*	(64.307)*	(0.319)**	(0.430)**	(0.432)**
Observations	8085	8085	8077	4243	4243	4240

Notes: Fixed effects regressions include facility fixed effects
Omitted illness category is chronic cases
Robust standard errors in parentheses
+ significant at 10%; * significant at 5%; ** significant at 1%

Table 4: Differential Side Payments for Acute Illnesses in High-bribe vs. Low-bribe Locations

Dependent variables	Wait		Log Wait		Side Payment Amount		Pay Side Payment		Log Side Payment		Side Payment as % of Total	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Log expenditure	-13.154	-13.249	-0.13	-0.13	25.824	25.832	0.066	0.066	0.266	0.266	0.038	0.038
	(7.969)+	(7.977)+	(0.054)*	(0.054)*	(10.382)*	(10.385)*	(0.012)**	(0.012)**	(0.085)**	(0.085)**	(0.008)**	(0.008)**
Acute	-12.203	2.533	-0.008	0.112	2.544	1.272	-0.017	-0.064	-0.068	-0.097	0.004	-0.003
	(7.151)+	(3.26)	(0.05)	(0.13)	(1.64)	(1.53)	(0.01)	(0.022)**	(0.14)	(0.41)	(0.01)	(0.01)
Acute*High-bribe	-14.15	-12.319	-0.221	-0.211	-30.651	-30.811	-0.15	-0.159	-0.156	-0.155	-0.048	-0.049
locations	(15.18)	(15.36)	(0.091)*	(0.092)*	(7.728)**	(7.579)**	(0.023)**	(0.023)**	(0.17)	(0.17)	(0.016)**	(0.017)**
Injury	-30.904	-1.792	-0.263	-0.34	0.664	-2.825	-0.008	-0.143	-0.292	-0.215	-0.03	-0.029
	(9.582)**	(5.89)	(0.100)**	(0.28)	(2.88)	(5.50)	(0.02)	(0.047)**	(0.18)	(0.73)	(0.010)**	(0.02)
Injury*High-bribe	-5.79	-3.649	-0.101	-0.104	-3.371	-3.617	-0.11	-0.122	0.31	0.31	-0.071	-0.071
locations	(22.64)	(22.61)	(0.15)	(0.15)	(12.00)	(11.89)	(0.034)**	(0.034)**	(0.22)	(0.22)	(0.019)**	(0.019)**
Interactions with provincial facilities	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Constant	156.913	157.54	4.254	4.253	-182.506	-182.562	-0.273	-0.275	0.282	0.281	-0.19	-0.19
	(64.956)*	(65.020)*	(0.432)**	(0.432)**	(84.784)*	(84.803)*	(0.100)**	(0.100)**	(0.73)	(0.73)	(0.065)**	(0.065)**
Observations	7725	7725	4036	4036	7475	7475	7497	7497	1402	1402	5795	5795

Notes: "Pay side payment" is an indicator for paying a positive amount. All regressions include facility fixed effects
Additional interactions with provincial facilities include interaction terms acute*indicator for provincial/district hospitals and injury*indicator for provincial/district hospitals
Other control variables: regressions in Columns 5-8 include official payment, columns 9-10 include log official payment
Omitted illness category is chronic cases
Robust standard errors in parentheses
+ significant at 10%; * significant at 5%; ** significant at 1%

Table 5: Differential Side Payments for Acute Illnesses in Central vs. Distant Locations

Dependent variables	Wait		Log Wait		Side Payment Amount		Pay Side Payment		Log Side Payment		Side Payment as % of Total	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Log expenditure	-13.36	-13.387	-0.134	-0.134	26.063	26.202	0.067	0.068	0.268	0.267	0.039	0.039
	(7.991)+	(7.994)+	(0.054)*	(0.054)*	(10.440)*	(10.421)*	(0.012)**	(0.012)**	(0.084)**	(0.085)**	(0.008)**	(0.008)**
Acute	-18.532	2.218	-0.115	0.109	-5.84	2.159	-0.067	-0.062	-0.172	-0.09	-0.013	-0.003
	(7.191)**	(3.23)	(0.045)*	(0.13)	(4.04)	(1.59)	(0.011)**	(0.022)**	(0.080)*	(0.41)	(0.01)	(0.01)
Acute*Central cities	14.16	21.119	0.299	0.42	-69.417	-58.625	-0.20	-0.143	-0.1	-0.057	-0.076	-0.057
	(21.39)	(23.01)	(0.23)	(0.237)+	(16.149)**	(17.875)**	(0.049)**	(0.050)**	(0.27)	(0.28)	(0.032)*	(0.034)+
Injury	-33.566	-1.899	-0.323	-0.338	0.361	-2.408	-0.055	-0.141	-0.083	-0.199	-0.063	-0.028
	(10.610)**	(5.97)	(0.076)**	(0.28)	(5.69)	(5.53)	(0.016)**	(0.047)**	(0.11)	(0.73)	(0.009)**	(0.02)
Injury*Central cities	3.26	7.248	0.62	0.684	-9.37	-8.633	0.04	0.083	0.301	0.214	0.044	0.077
	(29.27)	(33.03)	(0.364)+	(0.374)+	(41.17)	(41.68)	(0.12)	(0.12)	(0.36)	(0.36)	(0.06)	(0.06)
Acute*High-bribe		-14.176		-0.244		-25.615		-0.147		-0.151		-0.044
locations		(15.98)		(0.094)**		(8.360)**		(0.024)**		(0.17)		(0.017)*
Injury*High-bribe		-4.383		-0.14		-2.165		-0.124		0.294		-0.074
locations		(23.49)		(0.15)		(11.94)		(0.035)**		(0.22)		(0.020)**
Interactions with provincial facilities	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Constant	158.747	158.677	4.288	4.281	-184.311	-185.675	-0.276	-0.286	0.285	0.277	-0.194	-0.197
	(65.175)*	(65.165)*	(0.431)**	(0.431)**	(85.327)*	(85.112)*	(0.100)**	(0.100)**	(0.73)	(0.73)	(0.065)**	(0.065)**
Observations	7725	7725	4036	4036	7475	7475	7497	7497	1402	1402	5795	5795

Notes: "Pay side payment" is an indicator for paying a positive amount. All regressions include facility fixed effects
The central cities are Hanoi and HoChiMinh City
Additional interactions with provincial facilities include interaction terms acute*indicator for provincial/district hospitals and injury*indicator for provincial/district hospitals
Other control variables: regressions in Columns 5-8 include official payment, columns 9-10 include log official payment
Omitted illness category is chronic cases
Robust standard errors in parentheses
+ significant at 10%; * significant at 5%; ** significant at 1%

Table 6: Differential Side Payments for Acute Illnesses at Facilities of Different Audit Frequencies

Dependent variables	Wait		Log Wait		Side Payment Amount		Pay Side Payment		Log Side Payment	Side Payment as % of Total	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Log expenditure	-1.949	-2.146	0.101	0.072	-0.574	-1.806	0.027	0.011	-0.289	0.004	0.0002
	(4.51)	(4.49)	(0.30)	(0.31)	(2.95)	(2.74)	(0.05)	(0.04)	(0.58)	(0.02)	(0.02)
Acute	-1.356	-1.211	-0.254	-0.237	0.236	3.375	0.02	0.067	-0.156	0.014	0.032
	(2.96)	(3.05)	(0.29)	(0.31)	(2.29)	(1.791)+	(0.04)	(0.036)+	(1.17)	(0.02)	(0.017)+
Acute*No. of audits	0.20	0.198	0.047	0.046	-0.185	-0.133	-0.01	-0.008	-0.035	-0.002	-0.002
	(0.24)	(0.25)	(0.04)	(0.04)	(0.12)	(0.11)	(0.003)**	(0.003)**	(0.06)	(0.001)**	(0.001)**
Injury	-1.843	-4.983	-0.225	-0.422	-3.81	-0.587	-0.093	0.012	-0.419	-0.016	0.01
	(5.98)	(5.12)	(0.60)	(0.59)	(5.16)	(3.52)	(0.08)	(0.07)	(3.13)	(0.04)	(0.04)
Injury*No. of audits	-0.13	-0.143	-0.023	-0.014	-0.171	-0.089	-0.01	-0.006	0.035	-0.002	-0.002
	(0.37)	(0.38)	(0.08)	(0.08)	(0.26)	(0.23)	(0.01)	(0.004)	(0.41)	(0.002)	(0.002)
Acute*High-bribe		-1.618		-0.09		-27.091		-0.402			-0.133
locations		(5.27)		(0.40)		(7.835)**		(0.105)**			(0.041)**
Injury*High-bribe		15.509		0.907		-20.513		-0.613			-0.141
locations		(12.21)		(0.483)+		(16.62)		(0.152)**			(0.050)**
Constant	23.039	24.61	1.575	1.804	6.88	16.183	-0.017	0.095	4.045	0.027	0.057
	(35.40)	(35.23)	(2.38)	(2.41)	(21.60)	(19.98)	(0.36)	(0.34)	(4.90)	(0.16)	(0.16)
Observations	1064	1064	373	373	1051	1051	1051	1051	180	917	917

Notes: "Pay side payment" is an indicator for paying a positive amount. All regressions include facility fixed effects
Number of audits indicates the number of supervision/inspection visits each facility received during the 12 months prior to the interview, same period as the payment data
Regression of log side payment including high-bribe interaction terms (not reported) has highly colinear covariates and gives the same results as in column 9
Other control variables: regressions in Columns 5-8 include official payment, columns 9 include log official payment
Omitted illness category is chronic cases
Robust standard errors in parentheses
+ significant at 10%; * significant at 5%; ** significant at 1%