

Information and collective action in  
community-based monitoring of schools:  
Field and lab experimental evidence from Uganda\*

Abigail Barr, Frederick Mugisha, Pieter Serneels, and Andrew Zeitlin<sup>†</sup>

PRELIMINARY—NOT FOR CITATION  
August 2012

**Abstract**

Community-based monitoring of public services provides a possible solution to accountability problems when state oversight is limited. However, the mechanisms through which such policies can be effective are not well understood. Are community-monitoring interventions successful because they improve information alone, or do they also need to overcome collective action problems in local communities in order to be effective? We investigate this question by implementing a combined field and lab experiment in 100 Ugandan primary schools, which randomly assigns schools and their Management Committees (SMCs) either to standard community-based monitoring, to a participatory variation that addresses coordination problems, or to a control group. We find substantial impacts of the participatory treatment on pupil test scores as well as pupil and teacher absenteeism, while the standard treatment has small and insignificant effects, and we develop a test using randomization inference to show that differences in these joint outcomes between treated groups are statistically significant. Combining this evidence with SMC member behavior in laboratory games, we find evidence that improved collective action explains these differences. The results have implications for the design of community-based monitoring policies, and help to explain their variable effectiveness across contexts.

JEL codes: D78, I21, O15.

NEUDC program area and sub-area: Education (Education Delivery)

---

\*Barr: University of Nottingham; Mugisha: Economic Policy Research Centre and Ministry of Finance, Uganda; Serneels: University of East Anglia; Zeitlin: Georgetown University. We thank Lawrence Bategeka, Madina Guloba, and Ibrahim Kasirye of EPRC, Gregory Acar, Kees de Graaf, and Damalie Zalwango of SNV, and Judith Nakayamanna and Julius Odongo of World Vision. Seda Koymen, Sangamitra Ramachander, and especially Felix Schmieding provided superb research assistance. We are grateful to Paul Collier and to Stefan Dercon for helpful comments and encouragement.

<sup>†</sup>Corresponding author. Email: az332@georgetown.edu.

# 1 Introduction

Expansions in government supply and reductions in user fees have led to substantial improvements in primary education in many developing countries over the last 15 years. But these promising accomplishments have revealed new challenges, as the quality of education provided often remains poor. Perhaps nowhere is this more evident than in Uganda, where Universal Primary Education led to an increase from 3 million to more than 5 million pupils enrolled in primary school in its first year alone.<sup>1</sup> At the same time, learning lags behind neighboring Kenya and Tanzania, particularly in rural areas (Byamugisha and Senabulya 2005). At the heart of the problem, high rates of teacher absenteeism, estimated at 27 percent,<sup>2</sup> are often attributed to failures of accountability (Banerjee and Duflo 2006).<sup>3</sup>

To address such failures, policymakers have responded with three broad types of intervention (Bruns, Filmer and Patrinos 2011). Teacher incentive programs tie pay to (typically) centrally collected measures of performance. School-based management delegates control over resources to the local level. Information-for-accountability interventions, on the other hand, attempt to strengthen the short route of accountability by providing information alone—typically either by providing comparative information measures of performance, or by providing communities with training or tools to monitor their own schools.

This paper combines field and laboratory experimental evidence to study the impacts and mechanisms of community-based monitoring interventions in rural, government primary schools in Uganda. Such information-for-accountability interventions are attractive policy options for at least two reasons. Low economic and political costs contrast with pay-for-performance schemes, which have met with great resistance elsewhere.<sup>4</sup> Moreover, some

---

<sup>1</sup>These official enrollment gains are corroborated by Deininger (2003), who shows uses nationally representative household survey data to show that the fraction of children aged 6-12 attending primary school increased from 49 percent in 1992 to 73.5 in 1999. The government of Uganda responded by more than doubling the number of primary school teachers, and adding a further 88,000 classrooms in the 1996-2003 period alone (Kasirye 2009).

<sup>2</sup>These estimates, based on unannounced visits conducted by Chaudhury and coauthors (2006), exceed those authors' estimated rates of absenteeism for Bangladesh, Ecuador, India, Indonesia, and Peru.

<sup>3</sup>The World Bank (2004b) highlights two pathways of accountability. Along the 'long route', citizens hold politicians to account for service provision, and politicians respond by providing incentives or resources to service providers. A 'short route' of accountability, whereby citizens hold service providers to account directly, provides a potential alternative. In Uganda, the absence of a statistical association between PTA meeting frequencies and teacher absence rates (Chaudhury, Hammer, Kremer, Muralidharan and Rogers 2006) provides suggestive evidence of a failure of this short route.

<sup>4</sup>In Kenya, Chen and coauthors (2001) found that head teachers distorted records of teacher absenteeism when prizes were at stake, while in India, Banerjee and coauthors (2008) found non-compliance and sabotage made a health sector pay-for-performance

remarkable successes have demonstrated the potential of community-based monitoring: for example, Björkman and Svensson (2009) estimate substantial impacts of a ‘Citizen Report Card’ intervention on health clinic performance in Uganda, including a 33 percent reduction in child mortality.

However, the track record of community monitoring interventions is mixed. Olken (2007) finds no effects of increased community-based monitoring on corruption in road construction, in contrast with substantial impacts of top-down audits. Studying primary school performance in India, Banerjee and co-authors (2008) find that neither providing information about school performance, nor offering training to community members to assess pupil learning progress have an effect. De Laat, Kremer, and Vermeersch (2008), which is closest in context and policy content to our study, find no effect of combining teacher incentives with School Management Committee monitoring reports in Kenya (see Kremer and Holla 2009).

The reasons for these variable impacts are not well understood. Suggestive evidence indicates that failures to overcome collective action problems impedes the translation of information into improved school outcomes. Björkman and Svensson (2010) find that their intervention has greater effects in ethnically homogeneous communities. Drawing on evidence from other studies that ethnic heterogeneity is associated with low levels of public good provision,<sup>5</sup> they interpret this as evidence that program impacts depend on the “coordination of information and expectations”, and that ethnic fractionalization “adversely impact[s] collective action for improved service provision.” Banerjee and coauthors (2008) take a different approach, but arrive at a similar conclusion. Contrasting the non-impacts of two informational interventions with a third treatment, which also trained community members in the provision of supplemental literacy teaching, they attribute the effectiveness of this third treatment to having solved a collective action problem. However, neither paper can entirely exclude alternative explanations.<sup>6</sup>

To understand how community monitoring interventions work, and in so doing to inform policy design, we take a multifaceted experimental approach. This has two novel features.

---

scheme untenable.

<sup>5</sup>Alesina, Baqir, and Easterly (1999) and Alesina and La Ferrara (2000) provide theory and U.S. evidence on this relationship. Miguel and Gugerty (2005) demonstrate a relationship between local ethnic heterogeneity and parental contributions to schools in Kenya. Habyarimana and co-authors (2007) use a series of laboratory experiments in Kampala, Uganda to study the mechanisms by which ethnic diversity undermines local public good provision. We take this up in Section 2.

<sup>6</sup>In the case of Björkman and Svensson, it remains possible that co-ethnicity determines informational needs, rather than patterns of cooperation, while the literacy training of Banerjee and coauthors provides substantial human capital, relative to a pure informational intervention, which may explain its relative impacts. Challenges to the identification of the collective action mechanism are discussed in Section 2.

First, our field experiment implements two, related community-based monitoring interventions, each of which engages School Management Committees<sup>7</sup> in the use of a *school scorecard* monitoring tool. In the first of these, the *standard* treatment, SMC members were trained and supported in the use of a monitoring instrument that was developed in consultation with the Ministry of Education, District Education Offices, and NGO partners, and was informed by examples from other countries (World Bank 2004a). The second, *participatory* treatment differed only in that SMC members were engaged in a dialogue to design their own monitoring instrument, by defining objectives, roles, and indicators of progress.

We find strong effects of the participatory treatment across a range of school outcomes, but weaker and statistically insignificant effects of the standard approach. Using randomization inference to test for equality of impacts across three dimensions of outcome, we find evidence that the difference between the effects of these treatments is itself statistically significant. From a policy perspective, this provides evidence that small variations in intervention design have potentially large consequences, and the results demonstrate a high-return approach to strengthening accountability in the Ugandan context in particular. Because the two designs vary in the extent to which they help communities overcome collective action problems, this comparison also helps to speak to the broader question of mechanisms. In this sense, our field experiment is a variant on the class of mechanism experiments recently advocated by Ludwig, Kling, and Mullainathan (2011). By inducing experimental variation in the hypothesized mediating factor, collective action, we test whether activation of this mechanism is important to the success of the overall intervention. Because in a field setting it is difficult to vary this collective action component without also varying the information content of the monitoring instruments, a substantive focus of the paper is on identifying the collective action mechanism in particular. This motivates a second feature of our design.

The second novel feature of the experimental design helps us to identify the collective action mechanism. Immediately following training of SMC members, we conduct a post-intervention laboratory experiment in the field, in order to generate a direct measure of the differential effect of the treatments on community members' willingness to engage in collective action. Specifically, we play a binary public goods game with members of the SMC after the score card training in both our interventions, but not in the control group.<sup>8</sup> This game is based on the Dichotomous Voluntary Contribution

---

<sup>7</sup>As will be described in Section 3, the individuals charged with participating in our intervention were comprised largely of SMC members, but drew on additional members of the community for practical reasons. For expositional convenience, we refer to this implementing group as the SMC, but point out instances in which the distinction is likely to be important.

<sup>8</sup>We do not play the game with those in the control group because we are primarily

Mechanism of Marwell and Ames (1979), and follows the implementation of Cardenas and Jaramillo (2007).<sup>9</sup> Behavior in such public goods games often deviates from the self-interested Nash equilibrium, and variation in contribution rates provides a measure of SMC members' willingness to act collectively.

A small but growing number of studies are using laboratory games to obtain *outcome measures* following field experiments.<sup>10</sup> But while these lab measures have a theoretical relationship with further development outcomes, our paper is unique using the lab-type experiment as a measure of a mediating outcome, which is coupled with an analysis of final (educational) outcomes of interest.

We find that, in addition to out-performing the standard treatment in school outcomes, the participatory treatment induces higher contributions in laboratory public goods game. Taken together, these results confirm that participatory approaches can be effective and indicate that enhancing individuals' willingness to act collectively is at least part of the reason why this is the case. This throws some light on *how* community-based monitoring interventions work, and in so doing informs both the design of future interventions and their generalizability to other settings.<sup>11</sup> More broadly, these results provide evidence that, even within a single study context, variation in the design of development interventions has consequences for the efficacy of interventions that seek to address collective action problems.<sup>12</sup>

The remainder of the paper is structured as follows. Section 2 provides a theoretical framework for the analysis. Section 3 describes the experimental design, and results are provided in Section 4. Section 5 concludes.

---

interested in comparing mechanisms between the two treatments, and because we are wary of interference in the control group that might alter behavior.

<sup>9</sup>For related implementations of this game, see Attanasio, Pellerano, and Polania-Reyes (2009) and Barr and Serra (2011).

<sup>10</sup>Two recent papers have used experimental outcomes in studies of community-driven development schemes. Fearon, Humphreys, and Weinstein (2009) play a lab-type public goods game with treated and control communities in Liberia. Casey, Glennerster, and Miguel (2011) play what Harrison and List (2004) term a *framed field experiment* to measure impacts in Sierra Leone. Attanasio, Pellerano, and Polania (Attanasio, Pellerano and Polania-Reyes 2009) interpret a similar VCM game in terms of cooperation—and specifically, trust—as an outcome of a conditional cash transfer program in Colombia.

<sup>11</sup>Deaton (2009a, 2009b) has been perhaps the most forceful advocate that theories of mechanisms can inform questions of external validity.

<sup>12</sup>See King, Samii, and Snilstveit (2010) for a systematic review of community-driven development interventions that seek to promote social cohesion. Variable results across studies may be attributable to differences in design or context.

## 2 Mechanisms of community-based monitoring interventions

Community-based monitoring interventions are hypothesized to affect school outcomes by acting on one or both of two mechanisms: the information available to community members, and the community’s willingness to act collectively upon available information to improve school performance. Below, we describe the hypothesized interplay between policy interventions and these mechanisms. We then set out a framework for understanding alternative tests of these mechanisms, building on the measurement of mediating and moderating factors in the experiment.

### 2.1 Information and collective action mechanisms

By providing training or tools, community-based monitoring interventions may reduce the costs of gathering information about school processes and performance.<sup>13</sup> Richer and more credible information about school performance allows communities to better use their own resources—including social as well as pecuniary and in-kind rewards and sanctions—to incentivize teachers.<sup>14</sup> For example, if teachers are risk averse, incentives under an optimal contract become stronger as the noise with which their effort is observed decreases (Holmstrom and Milgrom 1991).<sup>15</sup>

The preceding argument presupposes a single principal, acting on behalf of the school’s clients and internalizing their costs and benefits. But when multiple clients are involved, classic public goods problems may arise. Both the gathering of information and the enforcement of rewards and sanctions require individuals to weigh private costs against public benefits. Moreover, either because of strategic complementarities in monitoring and sanctioning, or because individuals’ willingness to contribute public goods is contingent on their beliefs about the actions of others,<sup>16</sup> the effective collection and use of monitoring information may require solving a coordination problem among community members.

There is substantial evidence that ethnic heterogeneity in particular affects the provision of public goods, including evidence from Uganda. Hab-

---

<sup>13</sup>We distinguish between community monitoring and alternative forms of information-for-accountability interventions, including those that provide relative performance information across schools. The interventions studied here do not have such a ‘benchmarking’ component.

<sup>14</sup>Anecdotal evidence from Uganda suggests that the provision of meals and other in-kind gifts is used in part to reward good behavior by teachers.

<sup>15</sup>Alternatively, communities may be able to use this information to create pressure for improved allocations of school resources. This is potentially important given evidence that teachers distort these allocations (Pritchett and Filmer 1999). This would be expected to improve learning outcomes, but not necessarily teacher absenteeism.

<sup>16</sup>See, e.g., Fischbacher, Gächter, and Fehr (2001).

yarimana, Humphries, Posner, and Weinstein (2007) conduct a series of laboratory experiments with subjects from Kampala to test alternative mechanisms relating ethnic heterogeneity to low levels of public good provision. Significantly for our analysis, Habyarimana and coauthors argue that deleterious effects of heterogeneity are attributable to the importance of intra-ethnic sanctioning technologies for the enforcement of norms. In an education context, Miguel and Gugerty (2005) show that contributions to school funding in Kenya are negatively associated with ethnic diversity. This evidence suggests that—if the collective action mechanism is important—ethnically diverse communities will exhibit smaller gains from community-based monitoring interventions, a suggestion that is corroborated for the health sector by Björkman and Svensson (2010).

The *design* of community-based monitoring interventions themselves may also matter for whether they succeed in encouraging individuals to act in the collective interest.<sup>17</sup> More participatory interventions may, for example, provide community members with a context in which to signal both the value they place on the public good being created and their intention to contribute and, in turn, to update their beliefs about the values and intentions of others. The threat of social sanctions may help make such commitments credible. This suggests an alternative strategy to testing the collective action mechanism, by varying the extent to which interventions encourage collective action while, to the extent possible, holding other features of community monitoring constant.

## 2.2 Mediation vs. moderation—implications for experimental design

This paper follows the approach suggested above to test for a collective action mechanism in the effects of community-based monitoring. As detailed in the next section, we design two, related interventions which vary in the extent to which they activate this mechanism. Given inherent challenges in isolating this mechanism in the intervention design, we combine this field experiment with a direct, laboratory measure of willingness to contribute to public goods.

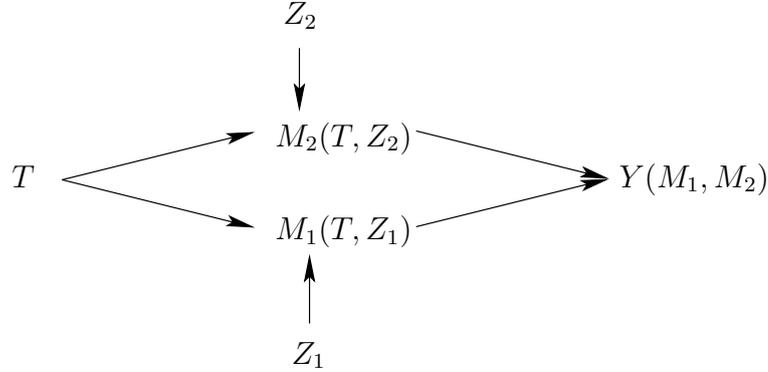
To clarify the role of field and laboratory experiments here, it is useful to distinguish between *mediating* and *moderating* (or *effect modifying*) factors.<sup>18</sup> Mediating factors are defined as intermediate outcomes that lie on the causal pathway between our experimental treatment and outcomes

---

<sup>17</sup>Banerjee and coauthors (2008) argue that a community-monitoring treatment which also trained community members in how to provide remedial reading education resolved such a coordination problem. As will be discussed below, it is difficult to distinguish the collective action effect *per se* from the impact of this added human capital.

<sup>18</sup>Testing of hypotheses about mediation, whether by structural or reduced-form methods, is a topic of active debate in statistics. See, inter alia, Baron and Kenny (1986); Sobel (2008); and Pearl (2009).

Figure 1: Mediation and moderation of policy impacts



Notes: Figure illustrates the impact on outcome  $Y$  of a treatment,  $T$ , when this effect is mediated through two intermediate channels,  $M_1$  and  $M_2$ . The effects of  $T$  on  $M_1$  and  $M_2$  are moderated by characteristics  $Z_1$  and  $Z_2$ , respectively.

of interest, while moderating factors—typically thought of as interaction effects—condition the effect size of the treatment under study.

Applying this distinction to the question at hand, community monitoring is presumed to impact school outcomes through improvements in the information available to hold teachers to account—call the state of available information mediating factor  $M_1$ . Our aim is to test the hypothesis that such interventions also succeed or fail by the level of collective action that they engender (mediating factor  $M_2$ ). As illustrated in Figure 1, let  $T$  stand for the policy treatment, reflecting either the status quo or any community-monitoring intervention in place, and let  $Y = Y(M_1, M_2)$  denote the final educational outcomes of interest.

One approach to testing the role of  $M_2$  is to design variations in  $T$ , say,  $T'$  and  $T''$ , which vary in their effects on  $M_2$ . This is the primary approach taken in this paper, and in Banerjee et al. (2008). In such approach it is valuable to have a direct measure of the mediator of interest,  $M_2$ —this is the role played by a post-intervention public good game in our study. Two problems arise without such a measure. If a treatment designed to activate  $M_2$  has no effect on  $Y$ , so that  $Y(T')$  and  $Y(T'')$  are the same, it is not possible to discern whether this is due to a failure to activate the mechanism (i.e., variation in  $T$  did not change  $M_2$ ), or whether the mechanism is truly unimportant for the outcome (i.e., there is no relationship between  $M_2$  and  $Y$ ). Moreover, if variation in  $T$  does have an impact on the outcome, so that the reduced form shows that  $Y(T') \neq Y(T'')$ , it may be impossible to entirely rule out the possibility that this operates through mechanisms other than  $M_2$ . Disentangling these mechanisms is a challenge

in our field experiment, as in Banerjee et al. (2008). But observed differences between  $M_2(T')$  and  $M_2(T'')$  can, at a minimum, help to substantiate the hypothesized mechanism. Induced changes in  $M_2$  are a necessary condition for the differences in  $Y$  to be ascribed to the mechanism of interest.

An alternative approach to testing for mediation is to make use of an effect moderator, such as the variables  $Z_1$  and  $Z_2$  in Figure 1. Suppose that factor  $Z_2$  modifies the effect of  $T$  on  $M_2$ , but does *not* moderate the effect of  $T$  on  $M_1$ , and suppose also that  $Z_2$  has no direct effect on  $Y$ . In this case, heterogeneity in the effect of  $T$  on  $Y$  along the dimension of  $Z_2$  provides evidence of the mediating channel  $M_2$ . Such evidence does not, strictly, require direct measurement of  $M_2$ . This is the approach taken by Björkman and Svensson (2010), who argue that heterogeneity in treatment effects according to ethnic fractionalization is evidence of coordination problems in community-based monitoring of health clinics. But it should be noted that the assumptions required here are strong:  $Z_2$  must not moderate the effect of  $T$  on  $M_1$ , must be linkable owing to prior empirical or theoretical work to  $M_2$ , and must have no direct effect on outcomes  $Y$ . A measure of  $M_2$  may again help to substantiate this test, by providing direct evidence that  $Z_2$  moderates the effect on the mechanism of interest.

These examples illustrate the importance of direct measurement of the hypothesized mediating variable. In what follows, we use a laboratory public goods game to measure our intervention variations' differential ability to encourage collective action.

### 3 Interventions, design, and data

#### 3.1 Scorecard interventions

The actual interventions evaluated in this project represent two variations on the notion of a *school scorecard* for community-based monitoring. School scorecards as a monitoring tool are an increasingly popular approach to what Bruns and coauthors call 'information-for-accountability' reform strategies (Bruns et al. 2011). Because information-only interventions involve 'low stakes' monitoring, they avoid some of the distortionary effects that have been observed in pay-for-performance schemes in some educational contexts (Glewwe, Ilias and Kremer 2010). Although the content of these scorecard interventions varies, a common approach uses them as a vehicle to involve community members in the gathering of information about school performance.

There are at least two channels through which such interventions may impact school outcomes. First, the information that they inject may be used by communities to hold schools to account, in a way that incentivizes improved performance. Even without the provision of external resources of explicit financial rewards, communities may be able to use nonpecuniary

benefits and social pressure to translate information into stronger incentives for service providers. Alternatively, when they foster dialogue, information interventions may facilitate cooperation within school communities, or between communities and service providers (Björkman and Svensson 2010). If their efforts are complementary—for example, if teachers only find it worthwhile to teach when parents help pupils with homework, and vice-versa—then this mutual cooperation can shift schools to a higher-performance equilibrium. Below, we describe two, related scorecard interventions that were designed to shed light on the mechanisms underlying successful community monitoring interventions.

### 3.1.1 Standard versus participatory scorecard

To test the importance of a participatory process as a means to enhance cooperation, we implemented two variants on the scorecard approach.

In schools allocated to the *standard scorecard*, we designed a scorecard over the course of a series of consultations with District and Ministry education officials, and project partners from the Netherlands Development Organisation (SNV) and World Vision and was piloted in schools outside of the study sample. This scorecard, which is presented in Appendix Figure B.2,<sup>19</sup> incorporates aspects of a range of existing monitoring tools, including those used by the District Inspectorate and as part of school-accountability programs run by SNV. The standard scorecard contains questions on themes of pupils’ involvement, provision for teachers, teacher presence and activities, materials and facilities, school finances, community involvement, health and wellbeing, and security and discipline. Under each theme, members of the SMC are provided with both quantitative indicators and a five-point scale to register their satisfaction with progress relative to the goals of the community.

By contrast, in schools allocated to the *participatory scorecard*, SMC members received the same training in the principles of monitoring and the development of objectives and indicators of progress. They then were led in the definition of their own goals and measures, starting from only a simple framework for a scorecard (see Appendix Figure B.1). The resulting participatory scorecard was thus distinct in each school in which it was used.

In spite of the loss of cross-school comparability within the participatory treatment arm, we hypothesized that the participatory scorecard might outperform the standard scorecard for one of two reasons. First, if problems facing schools even in similar locations are very different, such a ‘bespoke’ scorecard might better capture the informational needs of a particular school. Second, the act of defining goals and targets—the participatory design exercise itself—might facilitate the coordination of “expectations and

---

<sup>19</sup>Note that scorecards were translated into local languages for use in schools. Only the English prototype is presented here.

actions” (Björkman and Svensson 2010), and so encourage collective action.

### 3.1.2 Scorecard process

In an effort to isolate the coordinating effects of the participatory-design intervention, the remaining process of scorecard implementation was kept constant wherever possible across the two treatment arms. This process involved two steps: first, selection and training of individuals to participate in the use of the scorecard, and second, the collection and discussion of scorecard data each term.

The project timeline is provided in Appendix A. Selection and training of individuals to participate in the scorecard intervention was undertaken over the course of a three-day intervention in schools in October of 2009. These training meetings were led by Centre Coordinating Tutors (CCTs), who are staff of the Ministry of Education stationed in the districts for the purpose of providing in situ training to teachers. On the first day, a general meeting of the SMC, staff, and parents was called to explain the concept and to elect individuals to carry out the scorecard. To avoid the creation of parallel institutions, schools were strongly encouraged to nominate the existing members of the SMC unless there was an overriding reason not to do so, and the head teacher and SMC chair were given automatic roles. The scorecard committee consisted of a total of 12 individuals: three representatives each of teachers, parents, and management,<sup>20</sup> plus the head teacher and two pupils’ representatives (typically teachers acting as guidance counselors for male and female pupils, whose job would include solicitation of direct feedback from pupils). On the remaining two days, these elected participants would receive training in the underlying principles and the practical steps of this community monitoring intervention. In the case of the standard treatment, this involved learning the specific measures to be collected, whereas in the participatory scorecard, participants were involved in the design of the scorecard itself.

Once training was completed, the scorecard process was carried out in the same way in both treatment arms. In each term for the duration of the study, this process consisted of three steps. First, members of the scorecard committee would visit the school individually at least once during the term and complete their own copy of the scorecard. Second, at the end of the term, there would be a reconciliation process, in which scorecard committee members would meet, initially in small groups according to their roles, and subsequently as a whole, in order to agree upon a single set of scorecard

---

<sup>20</sup>Management representatives could be chosen from either the District Education Office or other centrally appointed representative on the SMC, or members of the ‘foundation body’ of the school. Foundation bodies are typically either a local church or mosque, or the local council; they play a continuing role in the management of the school and are represented on the SMC.

results for the term and to discuss specific goals and means for improvement in relation to this information. These meetings were facilitated by the CCTs. Third, the results of this ‘consensus scorecard’ would be disseminated, by sending it to the District Education Office and by discussing it at the next parent teacher association meeting.

### 3.2 Experimental design

We examine the impacts of these treatments in 100 rural primary schools. Four districts—Apac, Hoima, Iganga, and Kiboga—were chosen, spanning the regions of Uganda and capturing a range of the problems of poorly performing districts.<sup>21</sup> Schools were drawn from rural sub-counties only. For participation in the study, five sub-counties were chosen in each district, and five schools were chosen from within each sub-county. By sampling schools with probabilities proportional to size, we provide estimates that are representative of the school-going population in these areas.

Within this study population, schools were randomly allocated to treatments in order to evaluate program impacts. A total of 30 schools were assigned to each of the standard and participatory treatment arms, with the remaining 40 serving as a control group. This was done using a stratified random assignment, with sub-counties used as strata to balance the competing aims of comparability within strata and concerns over potential for contamination across study arms. Of five study schools per sub-county, two were assigned to control, and the remaining three were divided between the two treatments. Consequently, each district contains either seven or eight schools of each treatment type.<sup>22</sup>

### 3.3 Data

This paper makes use of three types of data.

First, we collected data on student learning achievements, together with survey-based and directly observed measures of school characteristics, at baseline and follow-up. To this end, we worked with officials from the Uganda National Examinations Board, who administered the National Assessment of Progress in Education (NAPE) exams at baseline to a representative sample of 20 pupils each in Primary 3 and Primary 6. These are the two years for which NAPE instruments are available; UNEB administers these exams to pupils from a representative sample of schools each year in

---

<sup>21</sup>It should be noted, however, that schools from Apac do not include many of the refugee-related issues that are pervasive farther north in the Northern Region.

<sup>22</sup>The total number of units in a given district receiving each treatment was selected at random, subject to the total number of units across districts. Similarly, within a given district, subcounties were first assigned to receive either more of the standard or more of the participatory scorecard (randomly, subject to the district quota), and then the randomization was conducted within that block.

order to provide indicators of educational progress to the Ministry. Because pupils in P6 had graduated primary school by the time of our follow-up survey, we focus analysis on the sample of Primary 3 pupils, who we tracked at follow-up.<sup>23</sup> The exams administered to each sampled student consisted of both a literacy and numeracy component. In addition, at follow-up we conducted unannounced visits in both treatment and control schools to measure absenteeism; these were conducted separately from survey and testing activities.

Table 1: School characteristics at baseline, by treatment assignment

	Control	Standard	Participatory	S-C	P-C
school size (pupils)	578.24 ( 334.30)	551.37 ( 220.02)	613.53 ( 299.22)	-26.87 ( 74.47)	35.29 ( 72.29)
pupil-teacher ratio	56.76 ( 24.97)	63.40 ( 25.60)	65.71 ( 25.40)	6.64 ( 6.40)	8.95 ( 6.27)
mean teacher absences	0.13 ( 0.08)	0.15 ( 0.11)	0.17 ( 0.10)	0.02 ( 0.02)	0.04 ( 0.02)
PLE pct Div. 1	0.01 ( 0.02)	0.01 ( 0.02)	0.02 ( 0.07)	0.00 ( 0.01)	0.01 ( 0.01)
PLE pct Div. 2	0.28 ( 0.20)	0.31 ( 0.20)	0.35 ( 0.22)	0.02 ( 0.06)	0.06 ( 0.05)
PLE pct pass	0.70 ( 0.17)	0.74 ( 0.17)	0.75 ( 0.17)	0.04 ( 0.05)	0.05 ( 0.05)
UNEB literacy z-score	0.10 ( 1.10)	-0.10 ( 0.94)	-0.04 ( 0.93)	-0.20 ( 0.24)	-0.14 ( 0.24)
UNEB numeracy z-score	-0.00 ( 0.99)	0.02 ( 1.03)	-0.01 ( 1.01)	0.02 ( 0.24)	-0.01 ( 0.24)

Notes: Columns (1)–(3) present means and standard deviations of variables, by treatment arm. Columns (4) and (5) present point estimates and standard errors for differences between standard scorecard and control and participatory scorecard and control, respectively. No such differences are significant at the 10% level or above. Teacher absences based on school records at baseline survey. Numeracy and literacy z-scores are school averages from standardized tests.

School-level characteristics from the baseline are presented in Table 1. These are broken down by treatment arm. This provides a test that the randomization ‘worked’, in the sense that it balanced observable characteristics across treatments. We observe no statistically significant differences across treatments here. Perhaps more substantially, it is notable that performance levels in the study schools are generally quite low: on average, only 1 percent of pupils achieves the highest division (Division 1) on the Primary Leaving Exam (PLE), and between 25 and 30 percent of pupils

<sup>23</sup> Analysis of impacts on test scores uses the balanced sample of pupils present at both baseline and follow-up. We discuss pupil attrition in relation to these estimates in Section 4.

who register for the PLE either fail it outright or do not complete the exam. Pupil-teacher ratios, while not out of line with national averages, are highly variable.

Second, because differences in informational content may also affect the outcomes of the participatory scorecard, we collected data on copies of both standard and participatory scorecards, as compiled by the District Education Offices. These monitoring data, which consist of scorecard marks in the case of the standard scorecard and questions designed in the case of the participatory scorecard, are discussed in Section 4.2.

Third, in order to provide a direct measure of the relative impacts of the two treatments on willingness to cooperate, we conducted a public goods game in both treatment arms, immediately following the introduction of the school scorecards. The sample for this game were the 12 individuals selected to participate in the scorecard training and subsequent exercise.

The specific public goods game played was a dichotomous Voluntary Contributions Mechanism (VCM) (Cardenas and Jaramillo H 2007). In this game, each subject is endowed with one token, which can either be allocated to a *private* or a *group* account. Tokens allocated to a private account return a value of UShs 5,000 (approximately USD 2.50 at the time) to the subject. On the other hand, tokens allocated to the group account return a value of UShs 1,000 to *all players* in the game. This was played as a one-shot, simultaneous-moves game, with all decisions recorded privately before the aggregate outcome was announced to the group and payoffs were made.<sup>24</sup>

The VCM game was chosen because it provides a laboratory analogue for the type of public goods problem inherent in all community-based monitoring interventions. The unique, dominant-strategy equilibrium for self-interested subjects in this game is for all subjects to keep their tokens in their private accounts and earn UShs 5,000. This equilibrium is Pareto dominated by an outcome in which all individuals allocate their token to the group account, in which case each individual earns UShs 24,000. Experimental economics has documented a strong tendency to deviate from the dominant-strategy equilibrium.<sup>25</sup>

Departures from the self-interested dominant strategies can be interpreted in two ways. Most obviously, these might reflect other-regarding preferences, such as altruism or inequality aversion. However, a growing body of evidence on *repeated* public goods games suggests that individuals are ‘conditional cooperators’ (Fischbacher, Gächter and Fehr 2001, Fischbacher and Gächter 2010): their propensity to contribute to public goods depends on their beliefs about others’ likelihood of doing so. Seen in this light, any differences in SMC members’ behavior between the standard and participa-

---

<sup>24</sup>Full details of protocols, including scripts, are available from the authors upon request.

<sup>25</sup>Cardenas and Carpenter (2008) summarize results from 9 VCM experiments in developing countries. Including continuous public good games, they find expected contribution rates of a between 30 and 80 percent of the initial endowment.

tory treatment arms may be attributable to changes in beliefs about their fellow members' willingness to contribute to public goods—both in the lab and in the field.

## 4 Results

In this section we report the impacts of the standard and participatory scorecard treatments on school and laboratory outcomes. A consistent pattern emerges from the school outcomes. Across a range of indicators, the participatory scorecard has substantial positive and statistically significant effects. Impacts of the standard scorecard are smaller, and consequently more difficult to distinguish statistically from zero in a small-scale experiment such as this. The picture that emerges from these results is one in which the participatory approach leads to higher effort levels from both the providers and clients of the schools, and improved learning outcomes result.

Because the two treatments will necessarily vary in information content as well as in the degree to which they foster cooperation, a fundamental challenge is to disentangle the two mechanisms. We present three types of evidence to support the argument that it is the coordinating process, and not the resulting information content, that is responsible for the relatively large impacts of the participatory treatment. First, we show that there are substantial differences in treatment effects even for outcomes—such as pupil and teacher absenteeism—that are very well measured in the standard scorecard. Since these are captured by all schools in the standard approach, but only by a subset of schools in the participatory approach, it seems unlikely that the superior performance of the participatory approach along these dimensions could be attributable to better information. Second, we document the content of the standard and participatory scorecards themselves. In so doing, we are able to show that they reflect a broadly similar set of problems and priorities. This suggests that comparatively small impacts of the standard scorecard are not a consequence of having focused on the wrong aspects of school processes. And third, we study effects of assignment into the participatory treatment on behavior in a public goods game. This provides direct evidence of a collective action mechanism at work.

### 4.1 Treatment effects on school outcomes

We present estimates of program impacts on three outcome measures: pupil test-score gains, pupil presence rates, and teacher presence rates.<sup>26</sup> Test-

---

<sup>26</sup>We also conduct similar analyses of pupil dropout and teacher retention, as discussed below and available upon request. Though average rates of both of these types of attrition are high in both cases, we find no effects of assignment to either treatment arm. For this reason, our results are interpretable as average treatment effects on those individuals who remain affiliated with the school between baseline and follow-up (Lee 2002).

score gains provide a measure of the ultimate objective of the interventions, the production of learning. Given the high rates of absenteeism documented earlier, increases in teacher presence capture a proximate objective of the intervention, while pupil presence impacts are open to interpretation either as a response to improvements in the quality of education or as an investment by parents.

#### 4.1.1 Impacts on learning

To estimate program impacts on learning gains, we make use of the cohort of pupils who were sampled to take the Primary 3 (P3) NAPE exam as part of the baseline survey. These pupils should in principle have been enrolled in P5 at the time of the follow-up survey, although prevalent grade repetition means that this is often not the case. The tracked cohort of pupils undertook the P3 exam in 2008 and the P6 exam in 2010.<sup>27</sup> We measure learning outcomes by pooling NAPE exams for literacy and numeracy. To evaluate learning impacts, we first convert the raw exam scores into subject and wave-specific z-scores, normalizing scores to have a mean of zero and a variance of one in untreated schools. This ensures comparability across years, since the P3 and P6 exams are marked on different scales and differ in difficulty from year to year.

An indication of the impact of treatment on learning outcomes can be seen from Figure 2. This figure displays the cumulative distribution of the z-scores in the follow-up tests, pooling numeracy and literacy scores, and grouping pupils by their treatment status. The distribution of scores under the participatory treatment in particular appears to be shifted to the right, reflecting the treatment effect on the middle of the distribution.

We test formally for learning impacts of the two interventions by estimating the following basic specification for the z-score of pupil  $i$  in subject  $j$  and school  $k$  at time  $t = 0, 1$ :

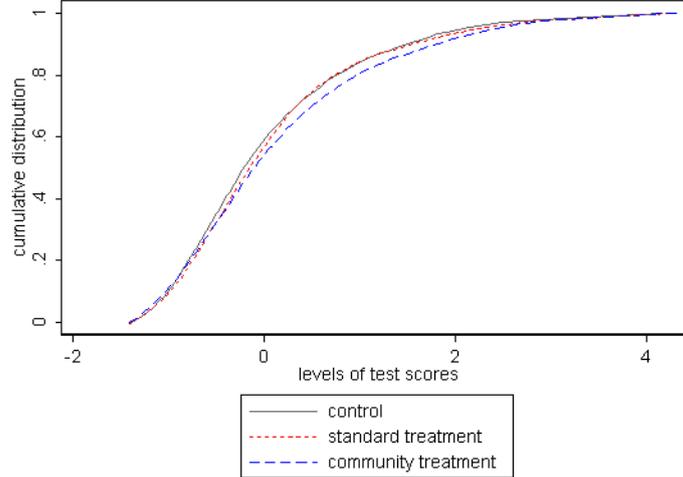
$$z_{ijkt} = \beta_0 + \beta_t t + \beta_P P_s + \beta_S S_s + \tau_P P_s t + \tau_S S_s t + \varepsilon_{ijkt} \quad (1)$$

where  $P_s, S_s$  are dummy variables taking a value of one if school  $s$  is in the participatory scorecard or standard scorecard groups, respectively. In this specification, the estimated treatment effect can be read off from the coefficients,  $\tau_P, \tau_S$ , on the interaction between the treatment assignment and the indicator for the follow-up exam (time  $t = 1$ ). The coefficients  $\beta_P, \beta_S$  capture any differences in average test scores across treatment arms in the baseline, prior to treatment.

---

<sup>27</sup>UNEB officials verified that the P6 exam would contain sufficient material accessible to P5 and even P4 pupils that it would be a sensitive instrument for measuring learning gains in our sample. This is borne out by the instrument's ability to detect treatment impacts.

Figure 2: Distribution of z-scores at follow-up, by treatment arm



Notes: Figure displays the cumulative distribution of z-scores among the tracked panel of pupils, using follow-up data only. Literacy and numeracy scores are pooled.

Table 2 presents estimates of equation (1), under alternative approaches to the error term  $\varepsilon_{ijkt}$ . In columns (1) and (2), we estimate a pooled OLS model, with column (2) adding controls for pupil characteristics. These specifications yield an estimated impact of the participatory scorecard of 0.19 and 0.22 standard deviations, which are statistically significant at the 10 and 5 percent levels, respectively. Estimated impacts of the standard scorecard are a little more than half of this magnitude and are statistically insignificant; however, given the considerable variation in exam performance, the differences between the two treatments are not statistically significant, as reported in the Wald test p-values below the table.<sup>28</sup> In columns (3) and (4), we use pupil- and pupil-exam fixed effects to address potential correlation between pupil or school characteristics and treatment assignment, and results are substantively unaffected. Note that, while the randomized assignment of schools to treatment should make this unnecessary in a sufficiently large sample and in the absence of selective attrition (an issue to which we return below), such a difference-in-differences specification provides an added degree of robustness.

It may be useful to give a sense of the magnitude of these impacts.

<sup>28</sup>The estimated coefficients on the assignment to participatory and standard scorecards ( $\beta_P, \beta_S$ ) are small in magnitude and statistically insignificant, allowing us to accept the hypothesis that the randomization effectively balanced these characteristics across treatment arms, leaving no pre-treatment differences between schools assigned to these programs and schools assigned to the control group.

Approximating the distribution of test scores with a normal distribution, the estimated impact of approximately 0.2 standard deviations would raise the median pupil 8 percentage points.

#### 4.1.2 Impacts on teacher and pupil attendance

We are interested in impacts on teacher and pupil attendance for several reasons. This is an essential input into the education production function: without attendance of teachers and pupils, there can be no learning. As documented by Banerjee and Duflo (2006), Chaudhury and coauthors (2006), and others, absenteeism problems are severe in many developing-countries, which makes this intermediate outcome important in and of itself as a proximate objective of policy.

Absenteeism outcomes can also help to shed light on the theoretical mechanisms at work in our study. The intuition for this is simple. Both pupil and teacher absenteeism are focuses of the standard scorecard, but feature only in a subset of the participatory scorecards designed by school communities. Consequently, if community monitoring works by providing means for the collection of information alone, one would expect the standard scorecards to have a higher impact, on average, on these indicators. But if enabling collective action problems to be resolved within the community is

Table 2: Program impacts on pupil learning outcomes

	(1)	(2)	(3)	(4)
	Pooled	Controls	Pupil FE	Pupil-exam FE
standard treatment $\times$ follow-up	0.0820 (0.10)	0.106 (0.12)	0.0786 (0.10)	0.0800 (0.10)
participatory treatment $\times$ follow-up	0.191* (0.10)	0.220** (0.11)	0.190* (0.10)	0.192* (0.10)
standard treatment	0.0259 (0.11)	0.00374 (0.13)		
participatory treatment	-0.0860 (0.13)	-0.114 (0.16)		
follow-up	0.529** (0.22)	0.230 (0.56)	0.340* (0.19)	-0.191 (0.18)
numeracy	0.0765** (0.03)	0.0809** (0.03)	0.0755** (0.03)	
Obs.	3512	3076	3512	3512
p-value	0.339	0.371	0.328	0.326

Notes: Dependent variable is standardized test z-score. Math and literacy test results pooled. Standard errors clustered at school level for all estimates. All specifications include strata-year controls. Additional controls for age and gender in column (2). p-value derived from Wald test of hypothesis that effect of treatments are equal.

Table 3: Program impacts on teacher and pupil absence

	(1)	(2)	(3)	(4)	(5)	(6)
	Teachers	Teachers	Teachers	Pupils	Pupils	Pupils
standard, $S$	0.0894 (0.06)	0.0914 (0.06)	0.115** (0.06)	0.0462 (0.05)	0.00669 (0.05)	0.0144 (0.05)
participatory, $P$	0.132** (0.06)	0.127** (0.06)	0.118** (0.06)	0.0896* (0.05)	0.0973** (0.05)	0.0992** (0.05)
baseline absence rate		-0.255* (0.15)	-0.0257 (0.23)			
yrs worked at school		0.0115* (0.01)	-0.00126 (0.01)			
ln baseline salary		-0.0163 (0.02)	0.00611 (0.03)			
female		-0.0147 (0.04)	-0.00595 (0.06)		0.0961*** (0.03)	0.0796* (0.04)
age					-0.0167 (0.01)	-0.0234 (0.02)
baseline mean z-score					0.0666*** (0.02)	0.0857*** (0.03)
$S \times$ base absence			-0.352 (0.32)			
$P \times$ base absence			-0.482 (0.31)			
$S \times$ yrs worked			0.0149 (0.02)			
$P \times$ yrs worked			0.0337*** (0.01)			
$S \times$ ln base salary			-0.313** (0.15)			
$P \times$ ln base salary			-0.0419 (0.04)			
$S \times$ female			-0.0642 (0.09)			0.0233 (0.09)
$P \times$ female			0.00424 (0.09)			0.0277 (0.07)
$S \times$ age						0.0187 (0.03)
$P \times$ age						0.00732 (0.02)
$S \times$ base mean z-score						-0.0569 (0.04)
$P \times$ base mean z-score						-0.0167 (0.04)
Obs	564	534	534	936	780	780
$H_1$ : p-value	0.512	0.574	0.148	0.356	0.0487	0.186
$H_2$ : p-value			0.0709			0.342
$H_3$ : p-value			0.0516			0.914

Notes: Linear probability model. Strata indicators included in all specifications. Dependent variable in columns (1)–(3) is indicator that teacher is present at endline unannounced visit; sample is all teachers employed at baseline and endline. Dependent variable in columns (4)–(6) is indicator that pupil is present during unannounced visit; sample is all pupils enrolled at baseline and endline. Wald test  $p$ -values presented for test of hypotheses that ( $H_1$ ) coefficients on standard and participatory treatments (and interactions where appropriate) are equal; ( $H_2$ ) impact of standard treatment is homogenous across observed pupil characteristics; and ( $H_3$ ) impact of participatory treatment is homogeneous across observed characteristics.

sufficiently important, then the participatory scorecard might outperform the standard scorecard in terms of absenteeism—even in spite of adding less information about absenteeism than the standard design.

We test impacts on teacher and pupil absenteeism using data from unannounced visits collected around the time of the follow-up study. In each case, we estimate as a base specification a linear probability model of the form

$$\Pr(y_{ikl} = 1) = \tau_S S_k + \tau_P P_k + \mu_l, \quad (2)$$

where  $y_{ikl}$  is an indicator variable taking a value of one if teacher or student  $i$  in school  $k$  and subcounty (strata)  $l$  is present on the morning of the unannounced visit,  $S_k$  and  $P_k$  are indicators for the standard and participatory treatments, and  $\mu_l$  is a strata-specific constant term (Bruhn and McKenzie 2009). To test for heterogeneity in treatment effects, we augment the specification in equation (2) with a vector of individual characteristics and treatment interactions.

Estimated impacts on the probability that teachers are present on the day of an unannounced visit show a substantial and statistically significant effect of the participatory scorecard in particular. These are presented in columns (1) to (3) of Table 3. We estimate treatment effects for the population of teachers employed by study schools at both baseline and followup, who were exposed to the treatment.<sup>29</sup> Teachers assigned to the participatory treatment are 13 percentage points more likely to be present in school on a randomly chosen day. This is a substantial gain, even when measured against the widespread absenteeism late in the school year: in control schools, only 51 percent of teachers who were employed at both baseline and endline are present on the day of the unannounced visit.<sup>30</sup> Estimated effects of the standard treatment are lower, at approximately 9 percentage points. This estimated effect cannot be distinguished statistically from either zero or from the participatory treatment, although our ability to reject equality of treatment impacts approaches statistical significance when observed dimensions of heterogeneity are included in the model.

We find some evidence that the effects of both the standard and participatory treatments on teacher presence are heterogeneous across observed characteristics of teachers.<sup>31</sup> The participatory intervention seems to out-

<sup>29</sup>Rates of teacher turnover in our data are high: 36 percent of teachers on the books at baseline are no longer employed by the same school two years later. Much of this is routine; firing and transfers are rarely used as a disciplinary device. We test for impacts of our treatments on the probability of continued employment. Point estimates are small and statistically insignificant.

<sup>30</sup>It should be noted that unannounced visits were conducted late in November, when absences are reported to become more frequent in advance of the PLE testing period. Consequently, this rate of teacher absence in control schools should not be taken as representative of the school year in general. However, the experimental results do show that this rate of absence is not an inevitable feature of that part of the school year.

<sup>31</sup>For each treatment, taken on its own, we are able to reject the hypothesis that the

perform the standard scorecard among more experienced and better paid teaching staff. For each additional year of teachers' tenure above the mean, the effect of the participatory treatment on teacher attendance increases by an additional three percentage points. On the other hand, the standard treatment is relatively *ineffective* among teachers with high salaries. A one standard deviation increase in log salary is associated with a decrease in the impact of the effect of the standard treatment by 46 percent—more than fully offsetting its effect.

Pupil attendance rates are valued both as a contributing factor to the learning outcomes already described, and as an outcome of policy interest in and of themselves. Over the long run, high attendance rates may contribute to a decrease in dropouts and improvements in grade progression.

In columns (4) to (6) of Table 3, we present impacts of the study interventions on pupil attendance. Estimated coefficients are from a linear probability model, with dependent variable equal to one if the pupil was present on the day of an unannounced visit to the school. Analogously to teachers, we estimate treatment effects for the population of pupils who were enrolled at both baseline and follow-up.<sup>32</sup> To address the concern that our treatments may differentially affect the probability of continued enrollment, we use a similar specification to test for impacts on enrollment probabilities. Point estimates of each treatment are very small (less than 4 percent in absolute value) and statistically indistinguishable both from zero and from each other.<sup>33</sup> Although we might hope for enrollment impacts from a policy perspective, these non-impacts simplify the analysis considerably. For example, the approach put forward by Lee (2002) and used by Kremer, Miguel, and Thornton (2009) to address selection collapses to ordinary least squares in the case where there is no selective attrition.

The estimated average treatment effect of the participatory treatment on attendance—ranging from 8 to 10 percent across specifications—is economically substantial and statistically significant. By contrast, the estimated effect of the standard treatment is smaller and less precisely estimated. In the absence of baseline data that would allow a difference-in-differences specification, we use a parsimonious set of control variables from the baseline data. This improvement in power allows us to reject the hypothesis that these two treatments have the same effect. In spite of the fact that female pupils are significantly more likely to attend school than boys (conditional

---

treatment effect is homogeneous across observed teacher characteristics at the 10 percent confidence level or better.

<sup>32</sup>Estimated rates of dropout and transfer are high, at 37 percent. Given our school-based sampling strategy, we are unable to distinguish between these two sources of discontinued enrollment.

<sup>33</sup>As a result, estimated impacts on the probability of attendance *without* conditioning on enrollment at follow-up are similar in magnitudes, and statistical significance is in fact increased in that specification, allowing us to reject equality of treatment effects without even the use of control variables to improve precision.

on being enrolled), and that attendance at follow-up is strongly correlated with test scores at baseline, we find no evidence of heterogeneity in impacts along either of these dimensions.

The estimates described above paint a consistent picture to the observed program effects on test scores and teacher presence. The effect of the participatory treatment on attendance is substantially larger than the standard treatment, and this difference is statistically significant in some specifications. Increases in pupil attendance may be part of a rational response on the part of parents to increases in teacher attendance, or this may reflect the community’s direct response to emphasis on the community’s role in improving education under the participatory scorecard. We take up the question of how the process and content of the participatory scorecard content relates to these observed patterns of impact below.

#### **4.1.3 Multiple comparisons and relative treatment impacts: A randomization inference approach**

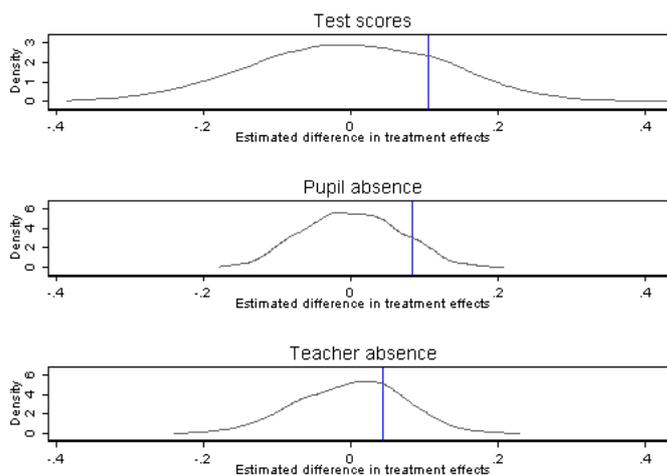
The comparison between treatment arms is central to the question of mechanisms here. Each of the outcome dimensions analyzed above—pupil learning, teacher presence, and pupil presence—potentially captures something of the same hypothesis: an improved ability to overcome collective action problems in the participatory treatment arm, relative to the standard arm. But although it is remarkable that there is a consistent pattern in relative treatment impacts across these dimensions, the small scale of the experiment leaves open the question of whether this pattern might arise by chance.

To address this question formally, we develop an intuitive approach to test whether this apparently consistent pattern of impacts across outcomes is statistically significant. This approach reflects three considerations. First, there is no a priori reason to privilege one dimension over another; each reflects an aspect of the process of interest. Second, the outcomes are different in their units of analysis (pupils and teachers) and their distributions. To address a similar issue in the evaluation of a community-driven development program, Casey, Glennerster, and Miguel (2011) adopt a mean-index approach (O’Brien 1984, Kling, Liebman and Katz 2007). In our case, where even the units of observation differ across outcome measures, we have little reason to believe that effect sizes will be the same, even if they capture the same theoretical mechanism. This makes an approach based on standardizing and pooling the outcome variables unappealing.

Instead, we build on the approach of ‘randomization inference’, first developed by Fisher (1935), and subsequently developed by others (see Rosenbaum 2002 for a discussion). The most common variant of this approach permutes the assignment of treatments at random in order to generate a distribution for a (scalar) test statistic under a null hypothesis of no treatment effect. We adopt this to estimate a two-sided test of the joint proba-

bility that we would observe as extreme an outcome as we observe for each of the dimensions under study, if the true model were one in which there were no difference in treatment effects.

Figure 3: Randomization inference: Simulated estimates of difference in treatment effects under null of equal effects



Notes: Marginal distributions of difference in estimated treatment effects for 1,000 draws under the null hypothesis of no difference. Blue line indicates the estimate in the true data.

To do so, we repeatedly reassign treatments, using the same stratified randomization procedure used in the actual experiment. We generate 1,000 such draws of the vector of treatment across schools, and for each of these draws, we estimate the *difference* in the treatment effects between the participatory and standard treatment arms, for each of the three dimensions of impact. This yields a  $3 \times \mathbf{1}$  vector,  $\tau_r$ , for synthetic treatments  $r = 1, \dots, 1000$ . The expected difference in treatment effects is known to be zero for each of these synthetic treatments, since they are orthogonal to the true treatment assignment in the data. The distribution of  $\tau_r$  reflects the small-sample variation and the non-independence across outcome dimensions. The resulting marginal distributions, together with the difference in treatments estimated from the true data, are presented in Figure 3. From these draws, we estimate the fraction of draws with a vector of treatment effects greater than the true estimate,  $|\tau_r| > |\tau_0|$  to be 0.048. This leads us to reject the hypothesis that the treatment effects are the same at the 5 percent level.

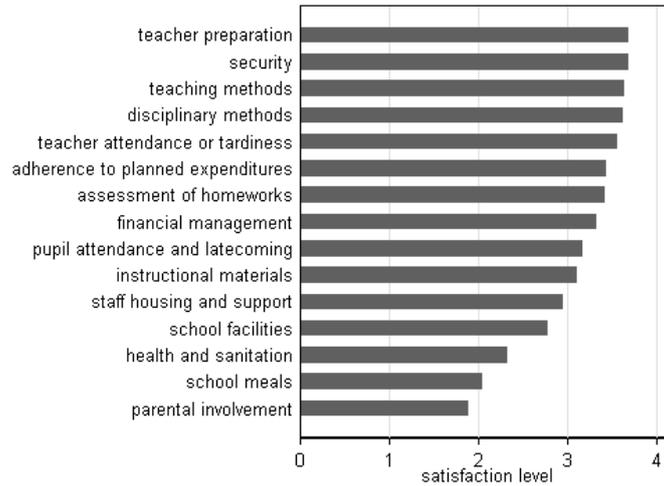
## 4.2 Scorecard content

The standard and participatory treatments differed in that SMC members in the standard treatment were trained in the use of a best-practice monitoring instrument, while SMC members in the participatory treatment were invited to design their own monitoring instrument in a participatory process that encouraged dialogue about aims for the school. While giving communities ownership of the process was considered to be important to overcoming collective action problems in the field experiment, this inevitably creates the possibility that there is variation in both the informational and the collective action mechanisms of the interventions. This poses a challenge for attribution of observed differences in outcomes across the interventions to the collective action mechanism in particular: the participatory instrument may also have better information, or information that is better tailored to the needs of the school in question. This potential informational advantage is difficult to address empirically because the content of the participatory scorecard may be selected for reasons that are associated with potential outcomes under alternative treatment assignments.

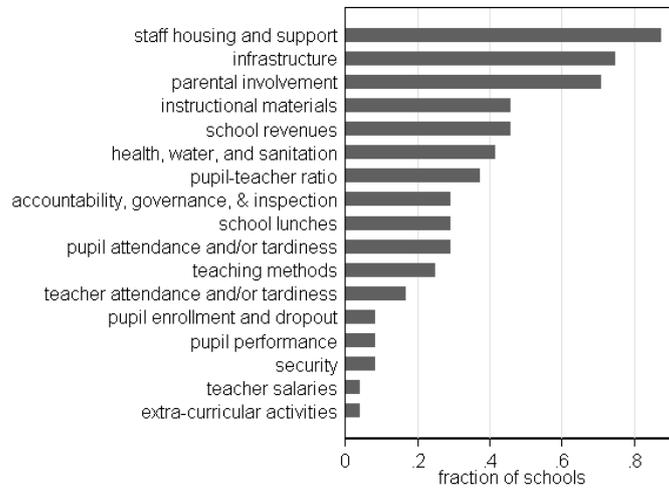
To address this challenge to isolating the role of the collective action mechanism, we use the content of the scorecards themselves—satisfaction levels reported in the standard treatment, and issues selected for inclusion in the participatory treatment. We use these data to show that, although the format of the two types of monitoring instrument is different, the content of each reflects a similar set of concerns about school performance. Moreover, we document in particular that pupil and teacher absenteeism—two domains in which the participatory approach out-performed the standard approach in the school—receive *less* attention in the participatory treatment. These observations cast doubt on the competing hypothesis that informational advantages in the participatory treatment were responsible for its relative performance.

SMC members in schools allocated to the standard treatment were provided with an opportunity to monitor progress and register their satisfaction across a range of thematic objectives and specific indicators, as illustrated in the scorecard design (Appendix Figure B.2). In Figure 4(a), we present average subjective satisfaction scores (on a 1–5 scale) for each of the broad scorecard dimensions. SMC members in the participatory treatment were tasked with selecting issues of concern to be considered in the scorecard exercise, and to identify specific indicators of progress along these dimensions. To do so, they were presented with a simple, blank format for a scorecard, as illustrated in Appendix Figure B.1. As part of the monitoring of the project, participatory scorecards were collected from 24 of the 30 schools in this treatment arm. These schools decided on an average of 5.75 issues each on their scorecards. The issues monitored by each school are summarized in Figure 4(b), which displays the fraction of schools including a given topic

Figure 4: Scorecard results



(a) Standard scorecard



(b) Participatory scorecard

Notes: Figure (a) gives the mean response across schools to each of the subjective assessments of thematic questions (from 1=“Very unsatisfactory” to 5=“Very good”). Figure (b) shows the percentage of schools assigned to the participatory scorecard treatment which elected to monitor an indicator of each issue. Administrative data available for all shared scorecard schools and for 26 of 30 schools in participatory treatment arm.

on their scorecard.

The priorities reflected in the participatory scorecard appear well reflected in the standard scorecard. Schools in the participatory treatment most frequently include staff support, infrastructure, parental involvement, and instructional materials, and school revenues as areas for monitoring. Each of these issues is captured in the standard scorecard, and indeed several are among the issues receiving the lowest subjective assessments on the standard scorecards. In spite of evidently high rates of teacher and pupil absenteeism, these issues feature less frequently in the participatory scorecards, and do not attract particularly low satisfaction ratings in the standard approach.<sup>34</sup>

The content of the participatory scorecards does not support the view that informational advantages in this treatment explain the relatively strong impacts of the participatory treatment on teacher and pupil absenteeism. Both the consequences and the root causes of absenteeism problems feature prominently in the standard scorecard, and are only included in a subset of the participatory schools' monitoring instruments. This lends credence to the hypothesis that the relative effectiveness of the participatory treatment stems from its success in fostering cooperation among stakeholders to address barriers to teacher performance.

### **4.3 Treatment effects on behavior in Voluntary Contributions Mechanism**

We hypothesize that impacts of the participatory treatment exceed those of the standard treatment primarily because of increased willingness to contribute to public goods, rather than differences in the information content of the scorecards. To corroborate this hypothesis, we would like to be able to measure and test for impacts on this hypothesized mediating outcome.

To provide such a measure we use the Voluntary Contributions Mechanism game, played immediately following the training of School Management Committees in their respective treatments. We chose to elicit an experimental measure of willingness to act collectively rather than survey-based reports of preferences, on the grounds that the use of financial incentives mitigates against experimenter-demand and Hawthorne effects (Orne 1969, Adair 1984).

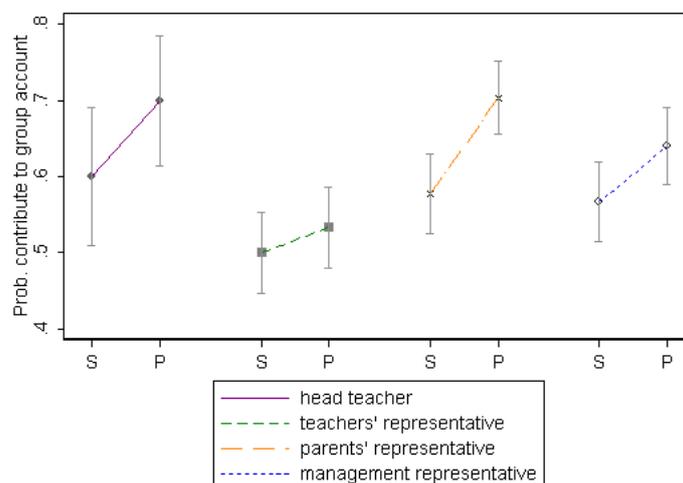
As described in Section 3, departures from the self-interested, dominant-strategy equilibrium in such public good games are typically ascribed either to (unconditional) social preferences, such as altruism, or to conditional co-

---

<sup>34</sup>This may partly reflect a design feature of the exercise. Concerned with creating tensions in the community, Ministry of Education facilitators emphasized the importance of finding constructive solutions instead of assigning blame when training SMC members. In this light, issues of staff housing and support can be seen as addressing root causes, rather than symptoms, of absenteeism problems.

operation. Accordingly, the participatory treatment might impact outcomes in the school and in the lab either by affecting preferences or by affecting beliefs about the willingness of others to contribute to public goods. Though we are unable to test between these, we find the latter more plausible. While a three-day training may make certain values more salient, we believe it unlikely that this would change preferences toward public goods sufficiently to have long-lasting effects on school outcomes. On the other hand, by offering SMC members an opportunity to publicly signal their values, the participatory treatment may have provided a shock to members' beliefs about each others' likely behavior.<sup>35</sup>

Figure 5: VCM contribution rates, by participant and treatment type



Notes: Figure displays estimates for average contribution rate to public account in the VCM, by stakeholder type and treatment arm (standard, *S*, or participatory, *P*).

Behavior in the VCM game is illustrated in Figure 5. Among each type of stakeholder, contribution rates to the shared account were greater under the participatory treatment. Ordinary teachers displayed lower rates of contribution under both treatments as compared to other stakeholder-types. Moreover, teachers relative response to the participatory intervention is smaller. In contrast, parents, the principle focus in local-accountability interventions, displayed a strong relative response to the participatory treatment.

Table 4 reports impacts of assignment to the participatory treatment

<sup>35</sup>In a laboratory context, Andreoni (1988) shows that deteriorating levels of cooperation in repeated public-goods games can be 'reset' by pausing the interaction. This confirms scope for manipulation of beliefs about cooperation.

Table 4: Parent contribution rates in VCM

	(1)	(2)	(3)
	all	all	parents
participatory treatment	0.0887*	0.162**	0.152**
	(0.05)	(0.08)	(0.06)
participatory × head teacher		-0.0607	
		(0.14)	
participatory × teacher		-0.141	
		(0.12)	
participatory × management		-0.0838	
		(0.12)	
participatory × ethnic share			0.878**
			(0.40)
head teacher	0.0183	0.0494	
	(0.07)	(0.11)	
teacher	-0.0969	-0.0247	
	(0.06)	(0.08)	
management	-0.0181	0.0247	
	(0.06)	(0.09)	
ethnic share	0.543***	0.540***	0.314
	(0.20)	(0.20)	(0.37)
Observations	550	550	166

Notes: Linear probability model. Dependent variable =1 if parent contributed to group account. Sample in columns (1) and (2) are all stakeholder types; sample in column (3) is parents only. Sub-county dummies and included in all specifications. Robust standard errors reported, clustered to allow non-independence at the session (school) level.

on contributions in the VCM. Note that because the VCM was conducted only in the two treatment arms, and not in the control group, we do not estimate impacts of the standard treatment relative to control. Our basic specification is a linear probability model of the form

$$\Pr(y_{ikl} = 1) = \tau_P P_{kl} + \beta X_{ikl} + \mu_l, \quad (3)$$

where the dependent variable  $y_{ikl}$  now indicates contribution of individual  $i$  in school  $k$  and strata  $l$  to the public account, and  $P_k$  is an indicator for assignment to the participatory treatment. Given the heterogeneity across subject types evident in Figure 5, we include a vector  $X_{ikl}$  of indicators for the subject’s role in the school. Both to directly test Björkman and Svensson’s (2010) hypothesis that ethnic heterogeneity impedes cooperation, as well as to improve power for this cluster-randomized design, we also include a session-level control for the proportion of parents in the school belonging to its largest ethnic group.<sup>36</sup> Finally, in all specifications, we include strata (sub-county) indicators  $\mu_l$ , reflecting the experimental design

<sup>36</sup>See, e.g., Bloom and coauthors (2005) for a discussion of the value of cluster-level controls for group-randomized designs in an educational setting. If the ethnicity variable

(Bruhn and McKenzie 2009), and we cluster standard errors to reflect the scope for non-independence within schools (or equivalently, experimental sessions).

The estimated average impact of the participatory treatment, across all participant types, is an 8 percentage point increase in the probability of contribution to the public good. Allowing for differential rates of response by participant type, however, we observe substantively and statistically stronger impact of the participatory treatment on parental contribution rates—the base category in column (2) of Table 4. The estimated 16 percentage point increase in parental contribution rates to the public good is statistically significant at the 5 percent level.<sup>37</sup>

Björkman and Svensson (2010) suggest that ethnic heterogeneity modifies the capacity of community monitoring interventions to encourage collective action in the school, and so that heterogeneity in impacts on service delivery outcomes provides evidence of the importance of collective action problems. The laboratory public goods game allows us to test this underlying assumption, without the auxiliary assumption that ethnic composition does not modify the effects of community monitoring through its informational channel. We do so by interacting the participatory treatment with an ethnic homogeneity measure for the sample of parents in column (3). We find not only that ethnic homogeneity is positively associated with contributions, but also that there is a significant, positive interaction between the participatory treatment and this measure of ethnic homogeneity.

## 5 Conclusion

This experiment has tested two variants on a ‘scorecard’ community-based management: a standard and a participatory approach, where the latter engaged School Management Committee members through the design of a school-specific monitoring approach.

Across a range of outcomes—pupil test scores, pupil presence, and teacher presence—we see a consistent story. The participatory design has substantial and statistically significant effects, while the standard approach is estimated to have smaller effects, and these effects are statistically significant. Although the small sample size of our pilot experiment has limited power, the differences between the two treatments are statistically significant for some dimensions and specifications, and a test by randomization inference rejects the null that, taken as a whole, the vector of treatment effects is equal. This gives reason to believe that the participatory approach has not

---

is excluded, the results pertaining to the participatory treatment are similar to those reported.

<sup>37</sup>Although point estimates suggest differences in both the level of contributions and, in column (2), their responsiveness to treatment, these differences are not statistically significant.

only has positive impacts, but may also outperform a standard design for such interventions.

There are at least two reasons why this may be the case. On the one hand, it is possible that the participatory design allowed information collected to be better tailored to the needs and preferences of school management. Alternatively, the participatory design may enhance the willingness of both teachers and parents to engage in collective action for improved service provision (Björkman and Svensson 2010).

We favor the second interpretation, for three reasons. First, the participatory scorecards rarely speak to issues that are not covered by the standard instrument. Second, if differences in information explain the differences in outcomes across treatments, we would not expect the participatory scorecard to outperform the standard scorecard on indicators such as teacher presence that are if anything *better* measured under the standard approach—but in fact we do. Third, a laboratory public goods game provides a direct test of a necessary condition for the collective action mechanism, namely that the participatory treatment increased individuals' willingness to act in the interest of the community, relative to the standard approach. This laboratory measure of the mediating mechanism also allows us to test the hypothesis that ethnic heterogeneity moderates the impact of the participatory treatment on collective action. Taken together, these findings provide suggestive evidence that the key feature of the participatory approach was that it better engaged the community in a process of discussing school goals, constraints, and progress.

These results have immediate implications for education policy in Uganda and similar contexts. Where accountability is low, and where test-based incentives may be expensive, information-for-accountability interventions provide a cost-effective alternative. The participatory scorecard approach evaluated in this project has strong effects at relatively low cost. More generally in the design of accountability programs, these results suggest that participatory engagement of the community—including the delegation of some authority over monitoring activities—may be essential to success.

## References

- Adair, J G**, “The Hawthorne effect: A reconsideration of the methodological artefact,” *Journal of Applied Psychology*, May 1984, *69* (2), 334–45.
- Alesina, Alberto and Eliana La Ferrara**, “Participation in heterogeneous communities,” *Quarterly Journal of Economics*, 2000, *115*, 847–904.
- , **Reza Baqir, and William Easterly**, “Public Goods and Ethnic Divisions,” *Quarterly Journal of Economics*, November 1999, *114* (4), 1243–1284.
- Andreoni, James**, “Why free ride? Strategies and learning in public goods experiments,” *Journal of Public Economics*, 1988, *37* (3), 291–304.
- Attanasio, Orazio, Luca Pellerano, and Sandra Polania-Reyes**, “Building trust: Conditional cash transfers and social capital,” *Fiscal Studies*, June 2009, *30* (2), 139–177.
- Banerjee, Abhijit and Esther Duflo**, “Addressing Absence,” *Journal of Economic Perspectives*, Winter 2006, *20* (1), 117–132.
- , **Rukmini Banerji, Esther Duflo, Rachel Glennerster, and Stuti Khemani**, “Pitfalls of participatory programs: Evidence from a randomized evaluation of education in India,” NBER Working Paper No. 14311 September 2008.
- Baron, R M and D A Kenny**, “The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations,” *Journal of Personality and Social Psychology*, 1986, *51*, 1173–1182.
- Björkman, Martina and Jakob Svensson**, “Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monitoring in Uganda,” *Quarterly Journal of Economics*, May 2009, *124* (2), 735–769.
- and —, “When is community-based monitoring effective? Evidence from a randomized experiment in primary health in Uganda,” *Journal of the European Economic Association*, April–May 2010, *8* (2–3), 571–581.
- Bloom, Howard S., Lashawn Richburg-Hayes, and Alison Rebeck Black**, “Using Covariates to Improve Precision: Empirical Guidance for Studies that Randomize Schools to Measure the Impacts of Educational Interventions,” MDRC Working Papers on Research Methodology November 2005.

- Bruhn, Miriam and David McKenzie**, “In Pursuit of Balance: Randomization in Practice in Development Field Experiments,” *American Economic Journal: Applied Economics*, October 2009, 1 (1), 200–232.
- Bruns, Barbara, Deon Filmer, and Harry Anthony Patrinos**, *Making schools work: New evidence on accountability reforms*, Washington, D.C.: The International Bank for Reconstruction and Development/The World Bank, 2011.
- Byamugisha, Albert and Frank Ssenabulya**, “The SAQMEQ II project in Uganda: A study of the conditions of schooling and the quality of education,” SAQMEQ Educational Policy Research Series, Uganda Working Report 2005.
- Cardenas, Juan-Camilo and Christian R Jaramillo H**, “Cooperation in large networks: An experiment,” Unpublished, Universidad de los Andes April 2007.
- Cardenas, Juan Camilo and Jeffrey Carpenter**, “Behavioural Development Economics: Lessons from Field Labs in the Developing World,” *Journal of Development Studies*, March 2008, 44 (3), 311–338.
- Casey, Katherine, Rachel Glennerster, and Edward Miguel**, “Reshaping institutions: Evidence on external aid and local collective action,” NBER Working Paper No. 17012 May 2011.
- Chaudhury, Nazmul, Jeffrey Hammer, Michael Kremer, Karthik Muralidharan, and F. Halsey Rogers**, “Missing in Action: Teacher and Health Worker Absence in Developing Countries,” *Journal of Economic Perspectives*, Winter 2006, 20 (1), 91–116.
- Chen, Daniel, Paul Glewwe, Michael Kremer, and Sylvie Moulin**, “Interim Report on a Preschool Intervention Program in Kenya,” Mimeo, Harvard University June 2001.
- de Laat, Joost, Michael Kremer, and Christel Vermeersch**, “Local Participation and Teacher Incentives: Evidence from a randomized experiment,” Working paper December 2008.
- Deaton, Angus**, “Instruments of development: Randomization in the tropics, and the search for the elusive keys to economic development,” Keynes Lecture, British Academy, October 9, 2008 January 2009.
- , “Understanding the mechanisms of economic development,” Unpublished, Princeton University August 2009.
- Deininger, Klaus**, “Does cost of schooling affect enrollment by the poor? Universal primary education in Uganda,” *Economics of Education Review*, 2003, 22, 291–305.

- Fearon, James, Macartan Humphreys, and Jeremy Weinstein**, “Can development aid contribute to social cohesion after civil war? Evidence from a field experiment in post-conflict Liberia,” *American Economic Review*, 2009, 99 (2), 287–291.
- Fischbacher, Urs and Simon Gächter**, “Social preferences, beliefs, and the dynamics of free riding in public goods experiments,” *American Economic Review*, 2010, 100 (1), 541–556.
- , — , and **Ernst Fehr**, “Are people conditionally cooperative? Evidence from a public goods experiment,” *Economics Letters*, June 2001, 71 (3), 397–404.
- Fisher, R A**, *The Design of Experiments*, Oliver and Boyd, 1935.
- Glewwe, Paul, Nauman Ilias, and Michael Kremer**, “Teacher incentives,” *American Economic Journal: Applied Economics*, July 2010, 2 (3), 205–227.
- Habyarimana, James, Macartan Humphreys, Daniel Posner, and Jeremy Weinstein**, “Why does ethnic diversity undermine public goods provision? An experimental approach,” *American Political Science Review*, November 2007, 101 (4), 709–725.
- Harrison, Glenn W and John A List**, “Field Experiments,” *Journal of Economic Literature*, December 2004, 42 (4), 1009–1055.
- Holmstrom, Bengt and Paul Milgrom**, “Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design,” *Journal of Law, Economics, and Organization*, January 1991, 7, 24–52.
- Kasirye, Ibrahim**, “Determinants of learning achievement in Uganda,” Paper presented at the annual conference of the Centre for the Study of African Economies, Oxford, 2009. February 2009.
- King, Elisabeth, Cyrus Samii, and Birte Snilstveit**, “Interventions to promote social cohesion in sub-Saharan Africa,” *Journal of Development Effectiveness*, 2010, 2 (3), 336–370.
- Kling, Jeffrey R, Jeffrey B Liebman, and Lawrence F Katz**, “Experimental analysis of neighborhood effects,” *Econometrica*, 2007, 75 (1), 83–119.
- Kremer, Michael and Alaka Holla**, “Improving education in the developing world: What have we learned from randomized evaluations?,” *Annual Review of Economics*, 2009, 1, 513–542.

- Kremer, Michael, Edward Miguel, and Rebecca Thornton**, “Incentives to learn,” *Review of Economics and Statistics*, February 2009, *91* (3), 437–456.
- Lee, David**, “Trimming for Bounds on Treatment Effects with Missing Outcomes,” NBER Technical Working Paper No. 277 June 2002.
- Ludwig, Jens, Jeffrey R Kling, and Sendhil Mullainathan**, “Mechanism experiments and policy evaluations,” *Journal of Economic Perspectives*, Summer 2011, *25* (3), 17–38.
- Marwell, Gerald and Ruth E Ames**, “Experiments on the provision of public goods. I. Resources, interest, group size, and the free-rider problem,” *American Journal of Sociology*, May 1979, *84* (6), 1335–1360.
- Miguel, Edward and Mary Kay Gugerty**, “Ethnic diversity, social sanctions, and public goods in Kenya,” *Journal of Public Economics*, December 2005, *89* (11–12), 2325–3468.
- O’Brien, Peter C**, “Procedures for comparing samples with multiple endpoints,” *Biometrika*, 1984, *40* (4), 1079–1087.
- Olken, Benjamin A.**, “Monitoring Corruption: Evidence from a Field Experiment in Indonesia,” *Journal of Political Economy*, 2007, *115* (2), 200–249.
- Orne, M T**, “Demand characteristics and the concept of quasi-controls,” in R Rosenthal and R Rosnow, eds., *Artifact in behavioral research*, New York: Academic Press, 1969.
- Pearl, Judea**, “Causal inference in statistics: an overview,” *Statistics Surveys*, 2009, *3*, 96–146.
- Pritchett, Lant and Deon Filmer**, “What education production functions really show: a positive theory of education expenditure,” *Economics of Education Review*, April 1999, *18* (2), 223–239.
- Rosenbaum, Paul R.**, *Observational studies*, New York: Springer-Verlag, 2002.
- Sobel, Michael E**, “Identification of Causal Parameters in Randomized Studies With Mediating Variables,” *Journal of Educational and Behavioral Statistics*, June 2008, *33* (2), 230–251.
- World Bank**, “Operational manual for Community Based Performance Monitoring,” Strategy for Poverty Alleviation Co-Ordination Office and Department of State for Finance and Economic Affairs, in collaboration with World Bank Social Development Department January 2004.

—, *World Development Report 2004: Making Services Work for Poor People*, Washington, D.C.: International Bank for Reconstruction and Development, 2004.

## Appendix A Project timeline

The project was carried out between the 2008 and 2010 school years, with the interventions in place in schools from the third term of 2009 to the third term of 2010.

**July 2008** Baseline study

**September 2009** Training of Centre Coordinating Tutors (CCTs) and District Education Office Staff

**October 2009** Training of School Management Committees by CCTs. First scorecard implemented in third term of 2009.

**January - November 2010** Scorecard implementation continues each term. Total of three visits by CCTs to facilitate termly 'consensus meetings'.

**November 2010** Follow-up survey and standardized testing

**November 2010** Unannounced visits to measure teacher and pupil attendance.

## Appendix B Standard and participatory scorecard formats

Figure B.1: Participatory design scorecard

Issue no.	Indicator	Symbol	Score	Reason
1				
2				
...				
10				

Figure B.2: Standard scorecard

<p>SCHOOL SCORECARD HEAD/TEACHER VERSION</p> <p>SCHOOL NAME: _____ SCHOOL ID: _____ DATE: _____</p> <p><b>Pupils' Involvement</b></p> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr><td>How many pupils are present in...</td><td>P1</td><td>SMILEY SCALE</td></tr> <tr><td></td><td>P2</td><td></td></tr> <tr><td></td><td>P3</td><td></td></tr> <tr><td></td><td>P4</td><td></td></tr> <tr><td></td><td>P5</td><td></td></tr> <tr><td></td><td>P6</td><td></td></tr> <tr><td></td><td>P7</td><td></td></tr> </table> <p>Are all the pupils in class at 9:00 AM? <input type="checkbox"/> Y/N</p> <p><b>Provision for teachers</b></p> <p>Have teachers received their last month's salaries by the 5th of this month? <input type="checkbox"/> Y/N</p> <p>Have teachers received meals or other resources in kind from the community this month? <input type="checkbox"/> Y/N</p> <p>Has anything been invested in construction or maintenance of staff? <input type="checkbox"/> Y/N</p> <p><b>Teaching activities</b></p> <p>Preparations. Do teachers have up-to-date... ...schemes of work? <input type="checkbox"/> Y/N ...lesson plans? <input type="checkbox"/> Y/N</p> <p>Presence. How many teachers are present by 8:30 AM on the day of your visit? <input type="checkbox"/> Y/N</p> <p>Methods. Observe one teacher for 1 hour. Which of the following teaching activities/methods did you observe? Reading aloud <input type="checkbox"/> Y/N Explanation or Discussion <input type="checkbox"/> Y/N Drill and Practice <input type="checkbox"/> Y/N Monitoring Student work <input type="checkbox"/> Y/N Monitoring Students <input type="checkbox"/> Y/N Resting <input type="checkbox"/> Y/N Projects <input type="checkbox"/> Y/N Are teachers using visual aids? <input type="checkbox"/> Y/N Are pupils asking questions? <input type="checkbox"/> Y/N Are pupils using textbooks? <input type="checkbox"/> Y/N Assessment. Has pupils' written work been assessed in the last week? <input type="checkbox"/> Y/N</p> <p><b>Materials and facilities</b></p> <p>Has the school acquired any textbooks or other learning aids since the end of last term? <input type="checkbox"/> Y/N</p> <p>Are classrooms well maintained? <input type="checkbox"/> Y/N</p>	How many pupils are present in...	P1	SMILEY SCALE		P2			P3			P4			P5			P6			P7		<p>SCHOOL SCORECARD HEAD/TEACHER VERSION</p> <p>SCHOOL NAME: _____ SCHOOL ID: _____ DATE: _____</p> <p><b>School finances</b></p> <p>Did you find the money received by the school listed publicly? <input type="checkbox"/> Y/N</p> <p>How much UPE money received since end of last term? <input type="checkbox"/> Y/N</p> <p>Does school have an approved budget? <input type="checkbox"/> Y/N</p> <p>How much money has been spent THIS TERM for purpose of... Instructional <input type="checkbox"/> Y/N Co-Curricular <input type="checkbox"/> Y/N Management <input type="checkbox"/> Y/N Administration <input type="checkbox"/> Y/N Community <input type="checkbox"/> Y/N</p> <p>Are receipts available for all expenditure? <input type="checkbox"/> Y/N</p> <p><b>Community involvement in the school</b></p> <p>Has the school held an Open Day this term? <input type="checkbox"/> Y/N</p> <p>Do homeworks have parent signatures? <input type="checkbox"/> Y/N</p> <p><b>Health and wellbeing</b></p> <p>Are latrines maintained well, with provision for drainage and daily cleaning/smoking? <input type="checkbox"/> Y/N</p> <p>Are there functional hand washing facilities near the latrines? <input type="checkbox"/> Y/N</p> <p>Are first aid facilities available? <input type="checkbox"/> Y/N</p> <p>Do all pupils have access to lunch? <input type="checkbox"/> Y/N</p> <p><b>Security and discipline</b></p> <p>Since the end of last term, has the school punished any students by use of... ... Corporal punishment? <input type="checkbox"/> Y/N ... Suspension? <input type="checkbox"/> Y/N ... Bullying? <input type="checkbox"/> Y/N ... Violence against girls? <input type="checkbox"/> Y/N ... drug or alcohol abuse? <input type="checkbox"/> Y/N</p> <p><b>Other Comments</b></p> <div style="border: 1px solid black; height: 40px; width: 100%;"></div>
How many pupils are present in...	P1	SMILEY SCALE																				
	P2																					
	P3																					
	P4																					
	P5																					
	P6																					
	P7																					

<p>How satisfactory is progress in pupil involvement in the school? <input type="checkbox"/> SMILEY SCALE</p> <p>How satisfactory is progress in provisions for staff? <input type="checkbox"/> SMILEY SCALE</p> <p>How well prepared do teachers seem to be? <input type="checkbox"/> SMILEY SCALE</p> <p>How satisfactory is progress in teachers' presence in this school? <input type="checkbox"/> SMILEY SCALE</p> <p>How satisfactory is progress in teaching methods? <input type="checkbox"/> SMILEY SCALE</p> <p>How satisfactory is progress in the assessment of homeworks? <input type="checkbox"/> SMILEY SCALE</p> <p>How satisfactory is progress in the supply of learning materials? <input type="checkbox"/> SMILEY SCALE</p> <p>How satisfactory is progress in the maintenance and upkeep of school facilities? <input type="checkbox"/> SMILEY SCALE</p>	<p>How satisfactory is progress in community involvement? <input type="checkbox"/> SMILEY SCALE</p> <p>How satisfactory is the management of school's finances? <input type="checkbox"/> SMILEY SCALE</p> <p>How satisfactory is the spending of school funds according to plan? <input type="checkbox"/> SMILEY SCALE</p> <p>How satisfactory are the maintenance and upkeep of sanitary and health facilities? <input type="checkbox"/> SMILEY SCALE</p> <p>How satisfactory is progress in the provision for student meals? <input type="checkbox"/> SMILEY SCALE</p> <p>How satisfactory are disciplinary methods in school? <input type="checkbox"/> SMILEY SCALE</p> <p>How satisfactory are steps taken in school to protect pupils' security? <input type="checkbox"/> SMILEY SCALE</p>
---	---

**SMILEY SCALE**

5. Very good  
4. Good  
3. Just OK  
2. Unsatisfactory  
1. Very unsatisfactory