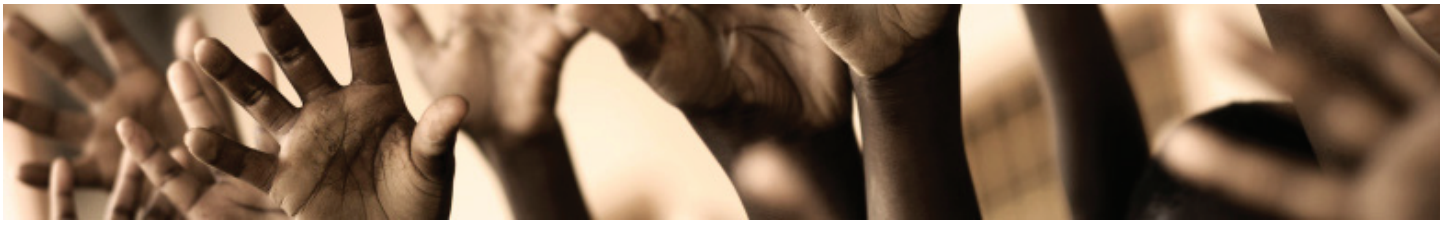
working paper
2012-14

Management and motivation in Ugandan primary schools: an impact evaluation report

Abigail Barr
Lawrence Bategeka
Madina Guloba
Ibrahim Kasirye
Frederick Mugisha
Pieter Serneels
Andrew Zeitlin

December 2012



partnership for economic policy pep

Management and Motivation in Ugandan Primary Schools: An impact evaluation report

Abstract

Among the various challenges that the Ugandan government is facing to improve educational outcomes and achieve Universal Primary Education (UPE) in the country, is the necessity to improve the "quality of education". Service delivery in education in Uganda has been proven to suffer, in great part, from the "weakness of accountability mechanisms between school administrators, teachers and the communities". In order to assist national decision-makers in solving these issues, a team of local researchers set out to test and assess the effectiveness of two types of community-based monitoring interventions in improving general educational outcomes, using methods of randomized controlled trials (RCTs) on a sample of 100 rural public primary schools in the country. This paper presents the main findings from this experimental impact evaluation project.

Keywords: Uganda, Universal primary education, education services, quality of education, community-based monitoring, accountability, school management and motivation, randomized controlled trial.
JEL codes: I21, I25, I28, I29

Authors

Lawrence Bategeka, Madina Guloba, Ibrahim Kasirye, and **Frederick Mugisha**
Economic Policy Research Centre, Kampala (Uganda)

Abigail Barr
University of Nottingham (United Kingdom)

Pieter Serneels
University of East Anglia (United Kingdom)

Andrew Zeitlin
Centre for the Study of African Economies, University of Oxford (United Kingdom)

This work was carried out with financial and scientific support from the Partnership for Economic Policy (PEP), which is financed by the Australian Agency for International Development (AusAID) and by the Government of Canada through the International Development Research Centre (IDRC) and the Canadian International Development Agency (CIDA), as well as the Government of the United Kingdom through the Department for International Development (DFID, or UK aid).

For more information, see www.pep-net.org

Executive Summary

We document the results of a randomized controlled trial that examined policies to improve functioning of School Management Committees (SMC) in rural, government primary schools in Uganda. The trial evaluated the impacts of two variations on a school monitoring scorecard, each of which was collected on a termly basis by SMC members. These treatments were designed to provide evidence not only on specific policy options for fostering 'bottom-up' accountability (World Bank 2004), but also to illuminate the importance of a participatory mechanism to achieve these effects.

Schools in the first treatment arm received training and support in a standardized scorecard, which incorporated best practices for simple indicators of pupil and teacher performance, teaching materials and facilities, and school governance. Schools in the second treatment arm received training in a participatory scorecard, which provided a forum for SMC members to develop indicators of dimensions of school performance that they valued themselves. Training was provided by Coordinating Centre Tutors (CCTs), who form part of the government educational staff resident in the study districts, and was overseen by the Netherlands Development Organisation (SNV) and World Vision (WV)-Uganda, working together with Economic Policy Research Centre (EPRC-Uganda) and Oxford staff. In addition, to test mechanisms of increased willingness to contribute to a public good, rather than rely exclusively on differences in the information content of scorecards, experimental laboratory games were played during implementation of the interventions.

Impacts of these alternative scorecard treatments were estimated using a sample of 100 schools from districts in each of Uganda's four regions: Apac, Hoima, Iganga, and Kiboga. To allow estimation of causal effects of the program, schools were randomly assigned to the standardized scorecard (30 schools), the participatory scorecard (30 schools), or control (40 schools). Randomization was stratified at sub-county level. The experimental procedure ensures that selective placement does not bias estimates program impact (see, e.g., Glewwe et al. 2004). Pupil and teacher absenteeism were measured at follow-up by use of unannounced visits to schools. Learning outcomes were measured by testing authorities from the Uganda National Examinations Board (UNEB), who administered tests from the National Assessment for Progress in Education (NAPE) to a representative sample of pupils at baseline and follow-up.

Results show statistically and economically significant effects of the participatory design scorecard, across a range of outcomes. The participatory design scorecard reduced pupil and teacher absenteeism by and 8.9 and 13.2 percent, respectively. The participatory scorecard had a

commensurate impact on pupil test scores of approximately 0.19 standard deviations; such an impact would increase a pupil from the 50th percentile to the 58th percentile of the distribution. Impacts of the standardized scorecard on these outcome measures are smaller and statistically indistinguishable from zero. Neither scorecard has a statistically significant impact on dropout rates or firing of teachers. With regard to the lab game results, findings indicate that across all participant types, the estimated average impact of the participatory treatment is an 8 percentage point increase in the probability of contribution to the public good. In addition, a parent's ethnicity played a positive impact on contributions to a public good-this reflected the degree of ownership in school.

These results suggest that the participatory design component of community-monitoring interventions may be important to their success. Delegation of this process appears to have fostered a stronger sense of ownership among school stakeholders. Given its low costs, such a participatory approach to community-based monitoring is a promising policy intervention for improving quality in UPE schools.

Acknowledgements

This project would not have been possible without the contribution of many people and institutions. We thank Gregory Acar, Kees de Graaf, and Damalie Zalwango of SNV, as well as Judith Nakamannya and Julius Odongo of World Vision for their contributions to the design and implementation of the intervention. We are grateful to Abigail Barr, Pieter Serneels, Martin Valdivia, Habiba Djebbari, Seda Koymen and Sangamitra Ramachander for their helpful comments, and especially Felix Schmieding for superb research assistance. Finally, we wish to thank Dr. Albert Byamugisha and the Ministry of Education and Sports, and in particular Sylvia Acana at the Uganda National Examinations Board, together with District Education Offices of Apac, Hoima, Iganga, and Kiboga districts for their thoughtful and indispensable collaboration with the project.

This document is an output from research funding by the UK Department for International Development (DFID) as part of the Improving Institutions for Growth (iiG), a research programme to study how to improve institutions for pro-poor growth in Africa and South-Asia, and from the Policy Impact Evaluation Research Initiative (PIERI) of the Partnership for Economic Policy (PEP). The views expressed are not necessarily those of DFID or PEP.

Table of content

1. Introduction

Since the advent of Universal Primary Education (UPE) in Uganda in 1997, there have been substantial gains in primary education. Enrolment gains have been the most notable. Using nationally representative data,[1]Deininger (2003) shows that the fraction of children aged 6-12 attending primary school increased from 49 percent in 1992 to 73.5 in 1999. These correspond to an increase from 3 million to more than 5 million pupils enrolled in the first year of UPE alone, according to official beginning-of-year enrolment statistics. Deininger (2003) shows that, these enrolment and attendance gains have been particularly strong among girls and among poorer households. While this rapid rate of increase in enrolment poses a natural challenge for physical resources in schools, the government of Uganda has 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).

In spite of these achievements, substantial challenges remain. This can be seen in pupil learning outcomes: according to the SACMEQ study of educational quality in Southern and Eastern Africa, Ugandan pupils in primary six lag behind average learning levels in these countries, including neighbouring Tanzania and Kenya (Byamugisha and Ssenabulya 2005). Low performance levels are particularly acute in rural areas. One possible explanation for performance problems going challenges can also be seen in rates of teacher absenteeism. Chaudhury et al. (2006) using unannounced visits to measure teacher absenteeism, find a teacher absenteeism rate of 19 percent in Ugandan primary schools. Such problems appear symptomatic of a failure of management and, in turn, accountability.

Policy interventions that seek to strengthen accountability can be thought of as operating through one of two channels (World Bank 2004). Under the 'long route' of accountability, citizens hold schools to account through political processes (e.g., voting), and government (both national and sub-national) manage these providers. The 'short route', by contrast, is direct: citizens may hold schools to account through direct interaction with the school. Parent-Teacher Associations (PTAs) and School Management Committees (SMCs) provide an institutional forum for this direct form of accountability. Potential strengths of the short route of accountability are several: the beneficiaries of a particular service have the strongest incentive to improve its performance, and they may also have the best access to information about the actual

[1]Deininger (2003) bases these estimates on the 1992 Uganda Integrated Household Survey (UIHS) and the 1999/2000 Uganda National Household Survey (UNHS).

performance of service providers..

Existing institutions of school management seem limited in their practical capacity to address these problems. Chaudhury et al. (2006) find no relationship between the frequency of PTA meetings and teacher attendance in Uganda. Baseline data collected for the present project suggest that parental participation in PTA meetings and other school activities is limited in scope, particularly outside of individuals holding positions of responsibility in the community (Kasirye 2010). While SMC members' attendance at meetings is uneven, with some key responsibilities-such as the co-signing of school accounts by the SMC Chair-seldom practiced (Guloba and Nyankori 2010).

Policymakers have intervened to strengthen the short route of accountability in various ways. One approach is to provide financial or other discretionary resources to local managers, which they can use to incentivize service providers. The track record of such interventions is mixed[2], although there is strong evidence that technocratic implementation of monetary incentives can improve effort by service providers (Duflo and Hanna 2006).

An alternative approach has focused on training clients in the monitoring of service providers-what Bruns et al. (2011) call an "information-for-accountability" approach. An example of this is the use of 'scorecards' to monitor the performance of service providers. In the health sector in Uganda, for example, Bjorkman and Svensson (2009) conduct a randomized, controlled trial, which demonstrates that the use of a 'citizen report card' to monitor primary health care providers can improve performance, resulting in (among other things) a 1.7 percentage point reduction in child mortality. Similar approaches have recently been employed in other countries and sectors, including in education.

In practice, recent intervention-for-accountability interventions in education have been varied in both design and results (see Bruns et al. 2011) for an overview). In a randomized, controlled trials in Madagascar, Lassibile et al. (2010) find impacts on school practices and pupil attendance and repetition from a bundled intervention that includes changes in school workflow as well as information, but no effects on teacher absence or pupil learning. In India, Muralidharan and Sundararaman (2010) evaluate the provision of professional diagnostic feedback, and again find changes in measured teacher behaviour (during classroom

[2]While SMCs with hiring and ring powers were effective in raising the performance of contract teachers in one experiment in Kenya (Duflo et al. 2009), experiments that gave discretionary resources to head teachers (Chen et al. 2001) or to School Management Committees (de Laat et al. 2008) for incentivizing regular teachers had no effect on outcomes such as teacher absence.

observations) but no impact on student learning. Banerjee et al. (2008) find no effect of either providing information to village education committees or of training school committees to gather information themselves on learning outcomes. Pandey et al. (2008) find that an information campaign that merely told communities about their responsibilities in school management had highly heterogeneous effects across states. And in Liberia, an information-only intervention that publicized reading assessment results and taught teachers to prepare quarterly report cards had only negligible effects (Piper and Korda 2010). Taking a different approach, Andrabi et al. (2009) find that providing information on relative performance to an entire educational market can cause bad schools to either improve or shut down.

Given this mixed evidence for the success of information-for-accountability, and the variety of policy designs piloted, comparatively little is known about two important and related issues:

1. How is the effectiveness of a community-monitoring intervention determined by design features of that intervention?
2. Through what mechanism do community-monitoring interventions work: by providing information, or by directly motivating stakeholders to contribute to the performance of the school?

The present project sheds some light on these questions. To do so, two variants on a school scorecard monitoring program were piloted and evaluated. In the first of these treatments, SMC members were trained in the use of a *standardized scorecard*, which was designed to reflect best practices of the Ministry of Education and Sports (MoES), Non-Governmental Organisation (NGO) partners, and experiences of other countries and sectors. In the second of these treatments, SMC members received the same training in monitoring principles, but were given the freedom to design their own scorecards-what we call herein the *participatory scorecard* approach. In both cases, SMC members collected data on the outcomes enumerated in the scorecard on a termly basis, and used these to set targets and plans for improvement. These interventions are described in detail in Section 2.

To test the efficacy of these interventions, this project implemented a randomized, controlled trial in 100 rural, primary schools. Schools in the sample were randomly assigned to one of the three treatment arms: standardized scorecard (30 schools), participatory scorecard (30 schools), or control (40 schools). Schools assigned to the control group were included in the baseline and follow-up surveys, but did not receive any intervention; these form a basis for comparison. The use of randomized assignment of schools to treatment arms is essential to the credibility of the analysis. Successful randomization ensures that any observed differences in

outcomes over the course of the trial are caused by the treatments themselves, since schools in all treatment arms will be comparable in terms of both observed and unobserved characteristics (Duflo et al. 2007). Details of the experimental design are provided in Section 3.

We document the implementation process, including the content of the participatory scorecards designed by SMC members, in Section 4. The analysis of treatment impacts is presented in Section 5. Section 6 concludes.

2. School 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 et al. (2011) call 'information-for-accountability' reform strategies. 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 education (Glewwe et al. 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 non-pecuniary benefits and social pressure to translate information into stronger incentives for service providers. Alternatively, when they foster dialogue information, interventions may facilitate coordination between service providers and communities (Bjorkman 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 coordinating effect can shift schools to higher-performance equilibrium. Below, we describe two, related scorecard interventions that were designed to shed light on the mechanisms underlying successful information-for-accountability interventions.

2.1 Scorecard process

In an effort to isolate the coordinating effects of the participatory-design intervention, the process of scorecard implementation and practice was kept constant 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.

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 Coordinating Centre Tutors (CCTs), who are staff of the MoES 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 PTA 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. The scorecard committee consisted of a total of 12 individuals: three representatives each of teachers, parents, and management[3], plus the head teacher and two pupils' representatives (typically the guidance counsellors for the school, whose job would include solicitation of direct feedback from boys and girls). On the remaining two days, these elected participants would receive training in the underlying principles and the practical steps of this information-for-accountability intervention (in the case of the participatory scorecard, they would also be involved in the design of the scorecard itself, as will be discussed below). Furthermore, an experimental behavioural game was played-Voluntary Contributions Mechanism (VCM)-with the elected scorecard committee members to access contribution to a public good.

Once training was finalised, the scorecard process was completed each term for the duration of the study. This process consisted of three steps. First, members of the scorecard committee would visit the school at least once during the term and complete their own/individual 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

[3] Management representatives could be chosen from either the DEO 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.

roles and come up with a group scorecard reflecting averaged scores of group members[4] and subsequently as a whole, in order to agree upon a single set of scorecard results[5] for the term. In addition, they would 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 a copy to the DEO's office and retain another copy which was discussed at the next PTA meeting.

2.2 Standard versus participatory scorecard

To test the importance of a participatory process as a means to coordinate expectations, we implemented two variants of the scorecard approach.

In schools allocated to the *participatory scorecard,* scorecard committee members received 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 A.1). The resulting participatory scorecard was thus distinct in each school in which it was used. Members were provided with a five-point scale to register their satisfaction on the indicators of progress they had identified to monitor throughout the course of the term.

By contrast, in schools allocated to the *standard scorecard,* we designed a scorecard over the course of a series of consultations with the District and MoES officials, and project partners from the Netherlands Development Organisation (SNV) and World Vision (WV), Uganda and was piloted in schools outside of the study sample. This scorecard, which is presented in Appendix Figure A.2[6], 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 scorecard committee are provided with both quantitative indicators and a five-point scale

[4] Implying that given that we have five groups (parents, teachers, SMC, pupil representatives and Head Teacher), five group scorecard reflecting averaged scores of individual scorecards would be formulated.

[5] The final 'consensus' scorecard would reflect the average score arising from group scorecards after negotiation.

[6] Note that scorecards were translated into local languages for use in schools. Only the English prototypes presented here.

(similar to the one under the participatory scorecard) to register their satisfaction with progress relative to the goals of the community.

In spite of the loss of cross-school comparability, 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 actions"[7].

There are many ways in which coordination problems might impede the progress of the school. The act of providing information and accountability itself has an element of coordination among SMC members. Evidence from elsewhere suggests that such coordination problems may be important: Banerjee et al. (2008) attribute the relative success of an intervention that provided training to volunteers in the provision of remedial education classes, when compared with a pure informational intervention, as arising from the fact that the training encouraged 'small-group action' that more easily overcame coordination problems. Alternatively, coordination between teachers and parents may be important efforts by each group to improve pupils' learning outcomes may be strategic complements.

3. Experimental design and data

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 poor-performing districts[8]. 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

[7] Bjorkman and Svensson (2010) emphasize this coordination problem as a factor explaining heterogeneous response to their intervention in health clinics in Uganda.

[8] 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. Ongoing work by Alderman et al. (2010) sheds light on educational constraints in such districts.

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 schools were divided between the two treatments. Consequently, each district contains either seven or eight schools of each treatment type[9].

Data for the project were collected at three points in time.

First, baseline data were collected in July of 2008. These included the administration of National Assessment of Progress in Education (NAPE) exams in numeracy and literacy by Uganda National Examinations Board (UNEB) personnel to a representative sample of 20 pupils each in Primary 3 and Primary 6. In addition, a school-level questionnaire collected basic administrative data, and individual-level questionnaires were administered to a representative sample of 5 teachers, 5 parents (selected from the parents of pupils sitting the P3 and P6 NAPE exams), and 5 SMC members, including the head teacher. Survey subjects also participated in a series of laboratory games, as documented by Barr and Zeitlin (2010, 2011).

School-level data 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.

[9]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, sub-counties 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.

Table 1: School characteristics at baseline, by treatment assignment

	(1) Control (C)	(2) Standard (S)	(3) Participatory (P)	(4) S-C	(5) P-C
School size (pupils)	**578.24**	**551.37**	**613.53**	**-26.87**	**35.29**
	(334.30)	(220.02)	(299.22)	(74.47)	(72.29)
Pupil-Teacher ratio	**56.76**	**63.40**	**65.71**	**6.64**	**8.95**
	(24.97)	(25.60)	(25.40)	(6.40)	(6.27)
Mean teacher absences	**0.13**	**0.15**	**0.17**	**0.02**	**0.04**
	(0.08)	(0.11)	(0.10)	(0.02)	(0.02)
UNEB PLE pct Div.1	**0.01**	**0.01**	**0.02**	**0.00**	**0.01**
	(0.02)	(0.02)	(0.07)	(0.01)	(0.01)
UNEB PLE pct Div.2	**0.28**	**0.31**	**0.35**	**0.02**	**0.06**
	(0.20)	(0.20)	(0.22)	(0.06)	(0.05)
UNEB PLE pct pass	**0.70**	**0.74**	**0.75**	**0.04**	**0.05**
	(0.17)	(0.17)	(0.17)	(0.05)	(0.05)
NAPE literacy z-score	**0.10**	**-0.10**	**-0.04**	**-0.20**	**-0.14**
	(1.10)	(0.94)	(0.93)	(0.24)	(0.24)
NAPE numeracy z-score	**0.00**	**0.02**	**-0.01**	**0.02**	**-0.01**
	(0.99)	(1.03)	(1.01)	(0.24)	(0.24)
School sample	40	30	30		

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. Overall unit of observation is the school.

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 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, data on the process of the intervention were collected during the training of SMC members, in October 2009. These data included basic characteristics of participants in the exercise, as well as the outcome of a behavioural game played at the conclusion of the training. Subsequently, the DEO's office compiled results of the first two rounds of scorecard data for monitoring purposes. 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 described in the Section 4.

Third, follow-up data were collected in November 2010. The follow-up data included abbreviated versions of the school and individual survey instruments used at baseline. UNEB-NAPE staff also conducted standardized testing of the tracked cohort of pupils who had sat the P3 exam at baseline in 2008[10]. As of 2010 they were expected in principle to be enrolled in P5; however, in practice their grades varied. UNEB administered the P6 exams to these pupils (this was a practical necessity, since NAPE does not test pupils at P5 or other levels). Since we are interested in comparing learning outcomes across treatment arms and not in measuring their absolute levels, this is problematic only to the extent that the test is so difficult that many P5 pupils would be 'bottom coded', receiving zero scores on the P6 exam, or that the exam would otherwise be insensitive to variations in pupil learning gains at the P5 level. However, UNEB officials verified that there was a sufficient range of questions on the P6 exam that a P5 pupil would be able to answer, such that the instrument would still be sensitive to differences in learning outcomes at that level[11]. And separately from the school visits on which testing and surveys were carried out, unannounced visits were undertaken to measure pupil and teacher absenteeism.

4 Implementation

4.1 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. A detailed timeline of project activities is provided in Table 2.

[10] To provide a baseline for subsequent studies, fresh cohorts of P3 and P6 were also sampled and tested. The sample was expanded to encompass an additional 20 schools not visited at baseline for this same purpose in a new sub-county.

[11] Scoring of the P5 pupils who sat the P6 NAPE exams was rescaled to suit P5 level-such that only P5 questions where considered when grading.

Table 2: Summary of project activities, 2008-2010

Activity	Period
Phase 1: Baseline	
1. Training of stakeholders	March 2008
2. Piloting of baseline survey instruments	May 2008
3. Fine tuning of survey instruments	June 2008
4. Training of UBoS field enumerators	July 2008
5. Baseline survey	July-August 2008
Phase 2: Interventions	
1. Piloting interventions	September, 2009
2. Training of DEOs and District School Inspectors on interventions & VCM game	September, 2009
3. Training of CCTs	September, 2009
4. CCTs' training and election of SMC on interventions	October-November 2009
5. Monitoring of intervention by EPRC, SNV, WV & CSAE	October 2009-November 2010
6. CCT Support scorecard committee members & school level monitoring	October 2009-November 2010
7. DEO receives scorecard	Every end of term from December 2009-December 2010
Phase 3: Endline	
1. Training of UBoS enumerators on endline survey instruments	October 2010
2. Endline survey	October 2010

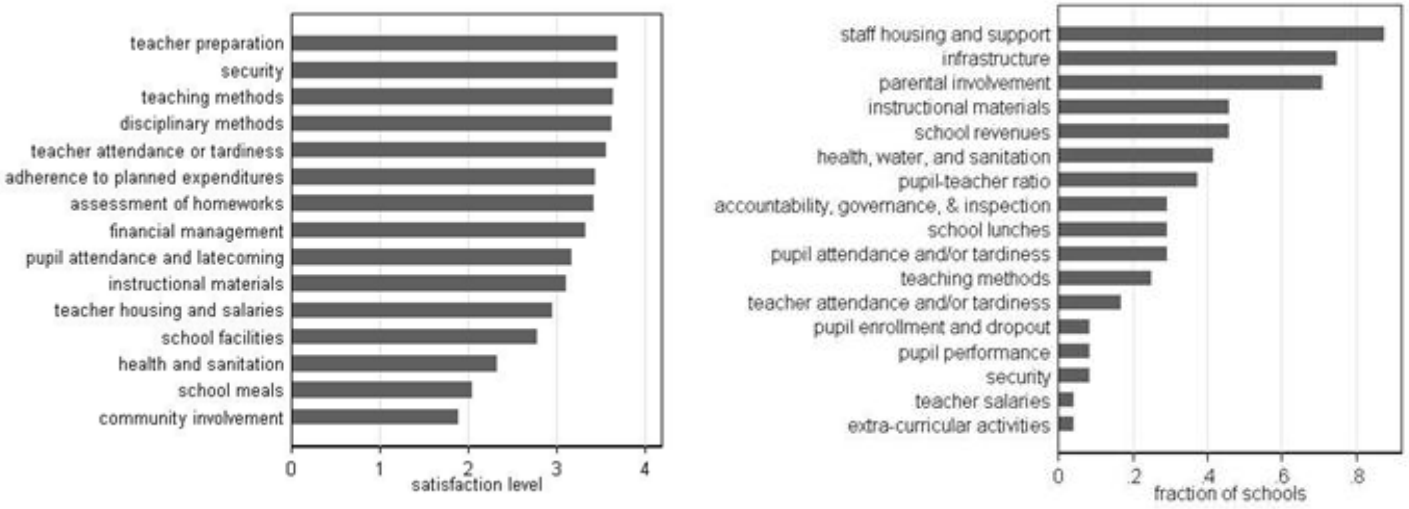
4.2 Scorecard contents

SMC members in schools allocated to the standard scorecard 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 A.2).

The results of the subjective assessments of progress from scorecards collected in the first two terms are presented in Figure 1. The results in Figure 1(a) seem to reflect a particular dissatisfaction with the state of joint involvement of the community and the school: the involvement of the community, the provision of school meals, and the improvement of school facilities are all rated among the worst areas and are all instances in which a substantial contribution is required from parents. By contrast, teachers-who it should be, remembered are also contributing to these scores-appear to be regarded as relatively well prepared, with only mild problems of attendance and teaching methods.

SMC members in schools assigned to the participatory scorecard were tasked with selecting issues of concern to be considered in the scorecard exercise, and to consider 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 A.1. As part of the monitoring of the project, these 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 1(b), which displays the fraction of schools including a given topic on their scorecard.

Figure 1: 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.

Three features of the qualitative choices of the participatory scorecard are striking.

First, it is evident that the voices of teachers are well reflected in the participatory scorecards. Teachers are represented on the SMC and consequently in the process of designing the scorecards. Teachers' concerns are reflected not only in the issue of staff housing, but also in the emphasis placed on holding *parents* accountable for supporting student learning. Given the clear emphasis on teachers' concerns, it is notable that teacher salaries are

rarely mentioned, although this may result from the perception that these are beyond the community's control.

Second, explicit discussion of teacher absenteeism was limited, but the root causes of absenteeism are widely mentioned in participatory scorecards. While teacher absences and late-coming are monitored in only 17 percent of schools, the lack of staff housing in such remote schools is typically cited as the dominant cause of this problem, and this is the most frequently included issue. This emphasis at addressing root causes of absenteeism may reflect the project leaders' emphasis on the importance of constructive framing-as opposed to "pointing fingers"-in the design of the participatory scorecard.

Third, both the standard and participatory scorecards reflect substantial concern over the ability of the school to finance running costs. This is reflected both in concerns over revenues (a topic that included the timely receipt of UPE funds, among other issues), as well as the provision of school lunches and the adequacy of facilities, salaries, and instructional materials.

Taken together, the participatory scorecards reflect a somewhat different interpretation of the problems facing rural primary schools than that which is typical of the economics literature. While there is evidence that teacher absences are considered a serious part of the problem, the scorecard content seems to reflect a view that teachers face substantial barriers to performing their duties. The most effective means to improving the quality of education may lie in mitigating these barriers, rather than in providing teachers with high-powered incentives and expecting them to resolve these issues themselves[12]. To foreshadow the results discussed in Section 5, it is possible that the relative effectiveness of the participatory scorecard stems from its success in coordinating the efforts of school stakeholders to address these obstacles.

5. Results

In this section, we report the main results of the project-the impacts of the standard and participatory scorecards on pupils, teachers, and management. A consistent pattern emerges from these findings. In addition, we report findings on the VCM experimental games played. Across a range of outcomes of interest, the participatory scorecard has substantial positive and

[12]This is consistent with the findings from the baseline laboratory experiments, which showed that-in an environment of low-powered incentives-teachers' *intrinsic motivation* is an important factor explaining their performance (Barr and Zeitlin 2010).

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.

5.1 Pupils

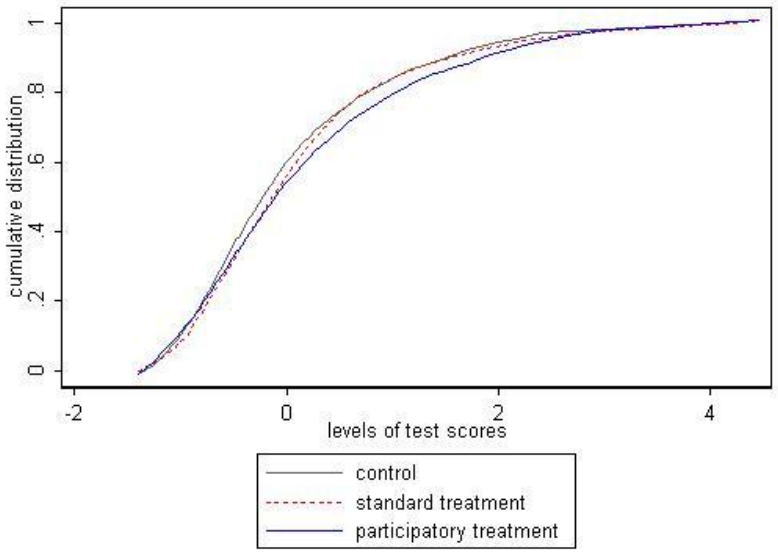
To estimate program impacts on pupils, we focus on the cohort of pupils who were sampled to take the Primary 3 (P3) exam as part of the baseline survey. These pupils should in principle have been enrolled in P5 at the time of the follow-up survey, although, as we will show below, prevalent grade repetition means that this is often not the case. Given the two-year interval between baseline and follow-up, pupils who were in P6 during the baseline survey had graduated by the time of the follow-up survey. Use of this panel of pupils who were tracked from P3 allows greater robustness and statistical precision through the use of a difference-in-differences empirical specification, as explained below.

5.1.1. Impacts on learning

We measure learning outcomes using NAPE exams for literacy and numeracy. The tracked cohort of pupils undertook the P3 exam in 2008 and the P6 exam in 2010. To evaluate learning impacts, we first convert the raw exam scores into z-scores, normalizing the scores to ensure that the scores have a mean of zero and a variance of one within the control group in each year. 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.

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.

Formally, we test forimpact of the two interventions on learning outcomes (intent-to-treat) 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, τ_P, τ_S, on the interaction between the treatment assignment and the indicator for the follow-up exam (time $t = 1$). The coefficients β_P, β_S capture any differences in average test scores across treatment arms in the baseline, prior to treatment.

Table 3 presents estimates of equation (1), under alternative approaches to the error term ε_{ijkt}.

Table 3: Program impacts on pupil learning outcomes

	(1)	(2)	(3)	(4)
	Pooled	**Controls**	**Pupil FE**	**Pupil-exam FE**
Standard treatment x follow-up	**0.0820**	**0.1060**	**0.0786**	**0.0800**
	(0.10)	(0.12)	(0.10)	(0.10)
Participatory treatment x follow-up	**1.191***	**0.220****	**0.190***	**0.192***
	(0.10)	(0.11)	(0.10)	(0.10)
Standard treatment	**0.026**	**0.004**		
	(0.11)	(0.13)		
Participatory treatment	**-0.086**	**-0.114**		
	(0.13)	(0.16)		
Follow-up	**0.529****	**0.230**	**0.340***	**-0.191**
	(0.22)	(0.56)	(0.19)	(0.18)
Obs.	**3,512**	**3,076**	**3,512**	**3,512**
p-value	0.339	0.371	0.328	0.326

Notes: Dependent variable is standardized test z-score. Numeracy 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.

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[13]. 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. Approximating the distribution of test scores with a normal distribution, the estimated impact of approximately 0.2 standard

[13]The estimated coefficients on the assignment to participatory and standard scorecards (β_P, β_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.

deviations would raise the median pupil 8 percentage points.

5.1.2. Impacts on enrolment and progression

We are interested in impacts of program assignment on the likelihood that the sampled P3 pupils at baseline remain in school at follow-up for two reasons. First, continued enrolment (and the successful progression of pupils through the classes) is a policy objective per se. Drop-out rates are strikingly high in this context. Among the representative sample of pupils who sat the P3 exam at baseline in control schools (i.e., in the absence of any policy intervention), only 63 percent remain enrolled in the same school at the follow-up study two years later[14].

Second, if the interventions considered in the present experiment affect dropout rates, then this would affect interpretation of the estimated impacts on test scores among pupils observed both at baseline and follow-up, as presented in Table 3. For instance it is theoretically possible that the participatory scorecard appears to positively affect learning when instead it causes selective dropout of individuals with low learning gains over the study duration.

To test for impacts on enrolment, we estimate a linear probability model of the form

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

where for instance y_{ikl} is an indicator variable taking a value of one if the pupil i in school k and sub-county (strata) l is enrolled at follow-up, S_k, P_k are indicators for the standard and participatory treatments, and μ_l is a strata-specific constant term (Bruhn and McKenzie 2009). Similarly, to analyze impacts on the class in which a pupil is enrolled, we regress class (taking values of 3 for pupils enrolled in Primary 3, 4 for pupils in Primary 4, etc.) on a set of treatment indicators and strata fixed effects.

Results for these outcomes are presented in Table 4. As reported in Column (1), we find no impact of either the standard or the participatory treatment on the probability of continued enrolment. This implies that although the participatory scorecard approach appears to have

[14]There are some apparent inconsistencies in the enrolment data provided by head teachers, as some of the pupils reportedly no longer enrolled did in fact participate in the follow-up exam. We report statistics treating such pupils as enrolled. This changes the enrolment rate in the control schools from 61 percent to 63 percent. It does not affect the substantive conclusions of this section.

been successful in boosting performance, it was not effective in addressing the problem of neither primary completion rates nor retention. Similarly, column (3) shows that there is no detectable difference in rates of progression across the treatments considered in the study.

Table 4: Program impacts on enrolment, participation in follow-up test, and grade progression

	(1)	(2)	(3)
	Enrolled	**Examined**	**Class**
Standard	**-0.0388**	**0.0327**	**0.0106**
	(0.03)	(0.05)	(0.07)
Participatory	**0.0155**	**0.0426**	**0.0414**
	(0.02)	(0.04)	(0.07)
Obs.	**1,071**	**1,071**	**976**
***p*-value**	0.128	0.843	0.679

Notes: All specifications include strata-specific constant terms (not shown). Standard errors clustered at school level. Sample in columns (1) and (2) is pupils who sat NAPE exam at baseline. Sample in column (3) is set of pupils who sat exam at baseline and are reported to be enrolled at follow-up.

In column (2) of Table 4, we demonstrate that the likelihood of sitting the follow-up exam is unaffected by treatment assignment. This is helpful from an analytical point of view, as under further assumptions it suggests that selective attrition is not driving the apparent test-score impacts reported in Table 3. For example, the approach put forward by Lee (2002) and used in Kremer et al. (2009) to address selection collapses to OLS in the case where there is no selective attrition.

5.1.3 Impacts on attendance

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 Table 5, 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. In columns (1) through (3), we test impacts on the probability of presence without conditioning on enrolment at follow-up; columns (4) through (6) repeat this exercise on the subset of pupils enrolled at follow-up.

The estimated impact of the participatory treatment on attendance-ranging from 8 to 10 percent across specifications-is economically substantial and statistically significant. This estimate is qualitatively unaffected by restricting the sample to those pupils who are enrolled at follow-up. By contrast, the estimated effect of the standard treatment is smaller and less precisely estimated. We are able to reject the hypothesis that these two treatments have the same effect in all but one of the specifications. In spite of the fact that female pupils are significantly more likely to attend school than boys, 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 similar picture to the observed program effects on test scores. The effect of the participatory treatment on attendance is substantially larger than the standard treatment, and this difference is statistically significant. Whether the increase in pupil attendance is a rational response on the part of parents to increases in teacher attendance, or whether this reflects the community's direct response to criticism of parental involvement in the scorecard exercise, will be revisited in light of impacts on teachers discussed below.

Table 5: Pupil Presence

	(1)	(2)	(3)	(4)	(5)	(6)
Standard	**-0.00235**	**0.00122**	**0.00741**	**0.0462**	**0.00669**	**0.0144**
	(0.05)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)
Participatory	**0.0809***	**0.0982****	**0.100****	**0.0896***	**0.0973****	**0.0992****
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Female		**0.0919*****	**0.0809****		**0.0961*****	**0.0796***
		(0.03)	(0.04)		(0.03)	(0.04)
Age		**-0.017**	**-0.0231**		**-0.0167**	**-0.0234**
		(0.01)	(0.02)		(0.01)	(0.02)
Baseline mean z-score		**0.0650****	**0.0844*****		**0.0666*****	**0.0857*****
		(0.02)	(0.03)		(0.02)	(0.03)
Standard x female			**0.00611**			**0.0233**
			(0.09)			(0.09)
Participatory x female			**0.0249**			**0.0277**
			(0.07)			(0.07)
Standard x age			**0.016**			**0.0187**
			(0.03)			(0.03)
Participatory x age			**0.00734**			**0.00732**
			(0.02)			(0.02)
Standard x baseline mean z-score			**-0.0552**			**-0.0569**
			(0.04)			(0.04)
Participatory x baseline mean z-score			**-0.0178**			**-0.0167**
			(0.04)			(0.04)
Obs.	**1,001**	**801**	**801**	**936**	**780**	**780**
H_1 : p-value	0.0734	0.0252		0.356	0.0487	
H_2 : p-value			0.407			0.342
H_3 : p-value			9.912			0.914

Notes: Linear probability dependent model. Dependent variable equals 1 if pupil was present in class during unannounced visit to school. Columns (1)-(3) present results for all pupils in P3 at baseline, while columns (4)-(6) present results for the subset of pupils who were enrolled at follow-up. p-values presented for Wald tests of the hypotheses that (H_1) standard and participatory treatments have same impact; (H_2) impact of standard treatment is homogenous across observed pupil characteristics; and (H_3) impact of participatory treatment is homogenous across observed characteristics.

5.2 Teachers

It is typically believed that information-for-accountability interventions operate through communities' ability to use this information to hold service providers accountable, possibly through the use of rewards or punishments outside of formal contracts. Alternatively, teachers may increase their effort levels under such interventions because and to the extent that the participatory nature of the intervention allows them to coordinate actions with community members-particularly important when the efforts of each group are complementary.

In this section, we test for impacts of the intervention on three out-come measures of intermediate interest: the probability that teachers are retained from baseline to follow-up; the probability that employed teachers are present in school on a given day; and the probability that present teachers are actually teaching at a given time. The results are presented in Table 6.

Program effects on teacher retention should be seen against a backdrop of dramatic turnover in employment among teachers. Of teachers in our control schools who were employed at baseline, 36 percent are no longer employed by the school a mere two years later. Teaching vacancies can take time to fill-especially for more senior positions. Thus, while increased probabilities of firing of malfeasant teachers is typically seen as evidence of improved accountability, this need not be the case: SMC members may not want to fire even underperforming teachers, for lack of an alternative. Moreover, even when SMCs are willing to re underperforming teachers, those same teachers may improve their e ort in response to this threat, so that no *equilibrium* increase in firing rates is observed.

Results on teacher retention appear to reflect this ambiguity. We observe no statistically significant effect of either treatment, relative to control, on the probability that a teacher employed at baseline remains with the school two years later[15].

By contrast, estimated effects on the probability that the still-employed teachers are present at the school on the day of an unannounced visit shows a substantial and statistically significant effect of the participatory score-card in particular. 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

[15]Because the point estimates have opposite signs, we are able to reject the hypothesis that the two treatments have the same effect, however. For reasons described above, interpretation of this result as a comparison in the relative effects on accountability is theoretically ambiguous.

baseline and endline are present on the day of the unannounced visit[16]. 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.

We find some evidence that the effects of both the standard and participatory treatments on teacher presence are heterogeneous across observed characteristics of teachers[17]. The effects of the participatory treatment are particularly strong for more senior teachers. That is, for each year of experience above the mean, the effect of the participatory treatment on teacher attendance increases by an additional three percent. We also find that 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 off setting its effect.

Finally, we find no effect of either intervention on the probability that a teacher present in school is found to be actually teaching at the time of the unannounced visit. These visits typically occurred near the outset of the school day, when 76 percent of teachers were found to be teaching in control schools. We also find no effect on the probability that a given teacher present in school has prepared a lesson plan for that day (results not shown).

To summarize, we find no effect of either treatment on teacher retention or on the activities of teachers found in school. However, we find substantial, positive impacts of the participatory treatment in particular on teacher presence. The participatory intervention seems to outperform the standard scorecard among more experienced and better paid teaching staff.

[16] 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 when teachers are posted to invigilate PLE exams in other schools and districts.

[17] For each treatment, taken on its own, we are able to reject the hypothesis that the treatment effect is homogeneous across observed teacher characteristics at the 10 percent confidence level or better.

Table 6: Program impacts on teacher retention, presence, and activities

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Employed			Present			Teaching		
Standard Scorecard	**0.0431**	**0.054**	**0.0516**	**0.0894**	**0.092**	**0.166****	**-0.113**	**-0.102**	**-0.0743**
	(0.03)	(0.03)	(0.03)	(0.06)	(0.06)	(0.06)	(0.08)	(0.08)	(0.08)
Participatory scorecard	**-0.0543**	**-0.0451**	**-0.046**	**0.132****	**0.129****	**0.119****	**-0.086**	**-0.0782**	**-0.0804**
	(0.04)	(0.03)	(0.04)	(0.06)	(0.06)	(0.06)	(0.07)	(0.07)	(0.08)
Baseline absence rate		**-0.392*****	**-0.315****		**-0.258***	**-0.0261**		**-0.225**	**-0.206**
		(0.08)	(0.13)		(0.15)	(0.22)		(0.17)	(0.28)
Years worked at school		**0.00717***	**0.0118***		**0.0114***	**-0.00115**		**0.00135**	**-0.00524**
		(0.00)	(0.01)		(0.01)	(0.01)		(0.01)	(0.02)
Log baseline salary		**-0.0132**	**-0.0121**		**-0.0164**	**0.00627**		**0.0259**	**0.0421**
		(0.01)	(0.02)		(0.02)	(0.03)		(0.02)	(0.04)
Standard x baseline absence rate			**-0.14**			**-0.36**			**-0.0821**
			(0.19)			(0.31)			(0.38)
Participatory x baseline absence rate			**-0.113**			**-0.488**			**0.00796**
			(0.22)			(0.31)			(0.46)
Standard x yrs worked at school			**-0.0118**			**0.0143**			**0.016**
			(0.01)			(0.02)			(0.02)
Participatory x yrs worked at school			**-0.00402**			**0.0334****			**0.00734**
			(0.01)			(0.01)			(0.02)
Standard x log baseline salary			**0.0091**			**-0.321****			**-0.214**
			(0.03)			(0.15)			(0.27)
Participatory x log baseline salary			**-0.00263**			**-0.0417**			**-0.0242**
			(0.02)			(0.04)			(0.04)
Obs.	**948**	**889**	**889**	**564**	**534**	**534**	**326**	**311**	**311**
H_1:p-value	0.00927	0.00614	0.111	0.512	0.565	0.131	0.758	0.796	0.922
H_2: p-value			0.518			0.0794			0.647
H_3: p-value			0.922			0.0253			0.923

Notes: Linear probability model. Strata controls included in all specifications. Dependent variable in columns (1)-(3) is indicator that teacher is still employed at endline; sample is all teachers employed at baseline. Dependent variable in columns (4)-(6) is indicator that teacher is present during unannounced visit; sample is all teachers employed at baseline and endline. Dependent variable in columns (7)-(9) is indicator that teacher is teaching at time of unannounced visit arrival; sample is all teachers employed at baseline and endline and present at unannounced visit. 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.

5.3 Management

Thus so far we have shown impacts of the participatory scorecard in particular on learning outcomes, and on teacher and pupil presence in schools. These final and intermediate outcomes may be brought about by changes in SMC and PTA behaviour in turn. To investigate this, we test for impacts of each treatment on two types of outcome: the conduct of SMC and PTA meetings, and the financial and in-kind contributions of parents to address school needs.

In Table 7 we estimate impacts on the frequency of and attendance at PTA and SMC meetings. To do so we use a difference-in-difference equation of the form in equation (1), where the unit of analysis is now the school-year and the outcome variables are defined appropriately. We find no significant effects of either program on any of these outcomes, either comparing them against the control schools or comparing them against one another.

Table 7: Program impacts on management activities

	(1)	(2)	(3)	(4)
	PTA meetings	**PTA attendance**	**SMC meetings**	**SMC attendance**
Standard x follow-up	**0.0155**	**21.92**	**-0.387**	**-0.0417**
	(0.53)	(21.46)	(0.38)	(0.08)
Participatory x follow-up	**0.71**	**26.00**	**-0.0089**	**-0.107**
	(0.52)	(21.14)	(0.37)	(0.08)
Follow-up	**-0.0271**	**-36.15*** **	**0.882*** **	**-0.0347**
	(0.34)	(13.83)	(0.24)	(0.05)
Standard	**-0.312**	**-22.46**	**0.0351**	**-0.0031**
	(0.39)	(15.51)	(0.27)	(0.06)
Participatory	**-0.247**	**-24.73**	**0.0026**	**0.0582**
	(0.37)	(15.06)	(0.27)	(0.06)
Obs.	**190**	**187**	**191**	**192**
$H_1 : p$ **-value**	0.219	0.859	0.345	0.458

Notes: Dependent variable in columns (1) and (3) is absolute number of meetings held in past year. Dependent variable in column (2) is number of parents attending most recent PTA meeting; dependent variable in column (4) is fraction of SMC members attending most recent SMC meeting. Strata indicators included in all specification. Wald p-value presented for test of hypothesis that treatment effects are equal, $\tau_S = \tau_P$.

Finally, we use a similar specification to test for impacts on community contributions and projects in the school. Given the attention paid to is-sues of staff housing, especially in the

participatory scorecard, one might expect to see the initiation of new projects along these lines. As shown in Table 8, we are unable to detect any impact on community contributions to infrastructure projects in this school.

Table 8: Program impacts on community contributions

	(1)	(2)	(3)	(4)	(5)	(6)
	anyprojects	**lnsmcvalue**	**teachany**	**lnteachspend**	**classany**	**lnclassspend**
Standard x follow-up	**-0.0858**	**1.838**	**0.053**	**1.130**	**0.0912**	**1.400**
	(0.17)	(2.10)	(0.10)	(1.44)	(0.08)	(1.17)
Participatory x follow-up	**0.0277**	**-0.0737**	**-0.0559**	**-0.885**	**0.0926**	**1.338**
	(0.16)	(2.12)	(0.11)	(1.45)	(0.08)	(1.19)
Follow-up	**-0.194***	**1.143**	**0.0471**	**0.382**	**-0.0591**	**-0.87**
	(0.11)	(1.42)	(0.07)	(0.98)	(0.06)	(0.80)
Standard	**0.0339**	**-2.082**	**0.0016**	**-0.134**	**-0.109**	**-1.577**
	(0.12)	(1.73)	(0.09)	(1.19)	(0.07)	(0.97)
Participatory	**-0.00907**	**-1.14**	**0.0013**	**-0.152**	**-0.108**	**-1.525**
	(0.12)	(1.75)	(0.09)	(1.2)	(0.07)	(0.98)
Obs.	**196**	**152**	**152**	**152**	**152**	**152**
$H_1 : p$ **-value**	0.52	0.393	0.327	0.191	0.988	0.96

Notes: Dependent variable, by column, is (1) **1**[any infrastructure projects]; (2) ln(total value community contributions+1); (3) **1**[any contribution to teacher accommodation]; (4) ln(value contribution to teacher accommodation+1); (5) **1**[any contribution to classroom construction or upgrading]; (6) ln(value contribution to classrooms+1). Strata indicators included in all specifications Wald p-value presented for test of hypothesis that treatment effects are equal, $\tau_S = \tau_P$.

5.4 Treatment effects on behaviour in Voluntary Contributions Mechanism

We hypothesize that impacts of the participatory treatment exceed those of the standard treatment primarily because of increased willingness to con-tribute 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.

In this project, we undertake a novel approach to testing mechanisms, by using a laboratory game played in the field. 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 included 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 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[18].

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[19].

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 co-operators' (Fischbacher et al. 2001, Fischbacher and Gachter 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' behaviour between the standard and participatory 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.

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'

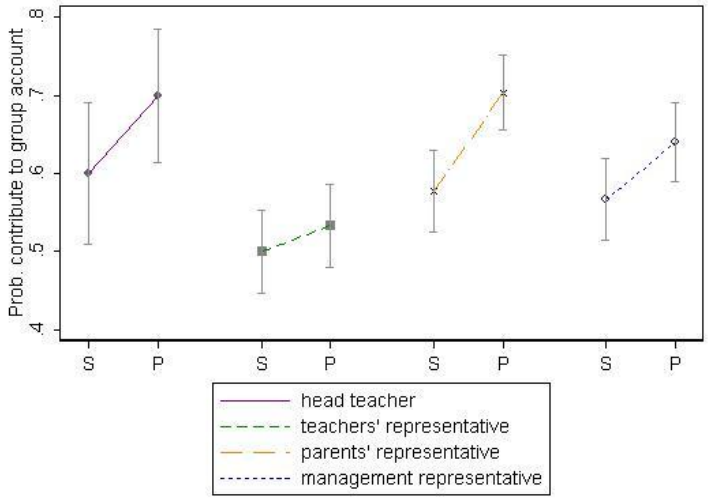
[18]Full details of protocols, including scripts, are available from the authors upon request

[19]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.

beliefs about each others' likely behaviour[20].

Behaviour in the VCM game is illustrated in Figure 3. 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.

Figure 3: VCM contribution rates, by participant and treatment type



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

Table 9 reports impacts of assignment to the participatory treatment 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

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

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

[20]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.

treatment. Given the heterogeneity across subject types evident in Figure 3, we include a vector X_{ikl} of indicators for the subject's role in the school. Both to directly test Bjorkman 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[21].Finally, in all specifications, we include strata (sub-county) indicators μ_l, reflecting the experimental design (Bruhn and McKenzie 2009), and we cluster standard errors to reflect the scope for non-independence within schools (or equivalently, experimental sessions).

Table 9: 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 x head teacher		**-0.0607**	
		(0.14)	
Participatory x teacher		**-0.141**	
		(0.12)	
Participatory x management		**-0.0838**	
		(0.12)	
Participatory x 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 are included in all specifications. Robust standard errors reported, clustered to allow non-independence at the session (school) level.

[21] See, e.g. Bloom et al. (2005) for a discussion of the value of cluster-level controls for group-randomized designs in an educational setting. If the ethnicity variable is excluded, the results pertaining to the participatory treatment are similar to those reported.

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 9. The estimated 16 percentage point increase in parental contribution rates to the public good is statistically significant at the 5 percent level[22].

Bjorkman 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.

5.5 Scorecard compliance

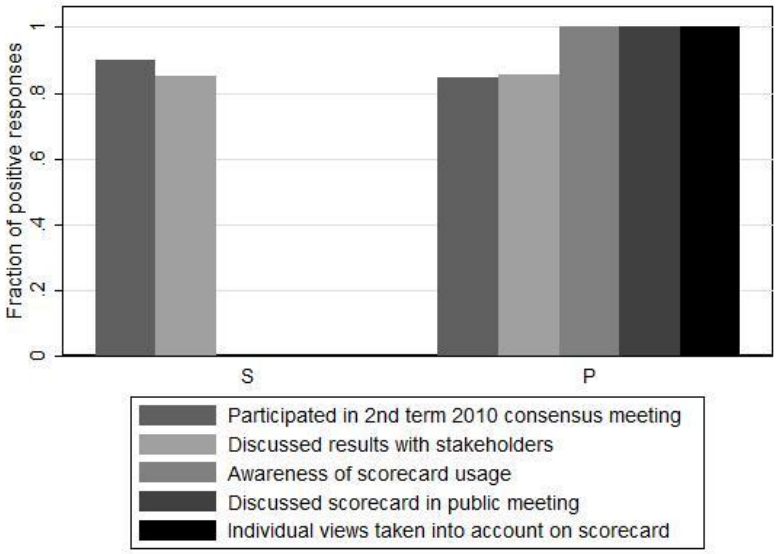
Given differences in school outcomes across treatments, and given the im-pact of the participatory treatment design on public goods contributions, a natural question is how this increased willingness to contribute to public goods was manifested in activities within the school that might have contributed to outcomes. To investigate this question, we use survey data on experiences with the intervention, focusing on levels of activity associated with the scorecard-which we loosely term as ‘compliance’ with the intended design-and on participants' satisfaction rates with the program.

In Figure 4, we report several measures of participation in the monitoring exercise, by treatment type. The sample for these responses is restricted to individuals who report being selected as a member of the scorecard implementation committee. Results suggest a striking contrast between the standard and participatory treatments. While roughly equal shares of

[22]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.

respondents report participating in the scorecard consensus meeting at the end of the most recent term and discussing scorecard results with other school stakeholders, there are marked differences in how the scorecards were used. Levels of awareness of how scorecards were being used, of having discussed scorecards in a public meeting, and of feeling that their individual views were taken into account in scorecard results were substantially lower in the standard approach.

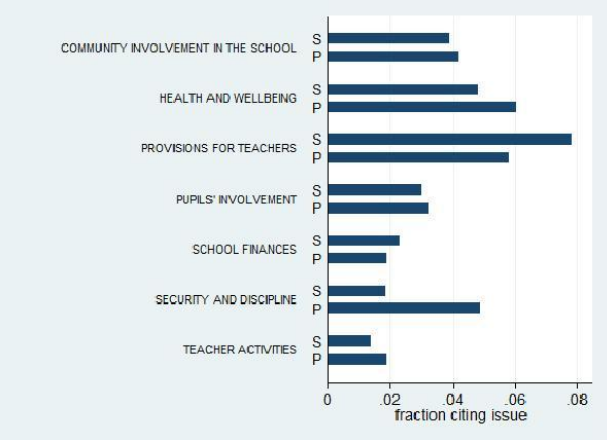
Figure 4: Scorecard-related activities, by treatment type



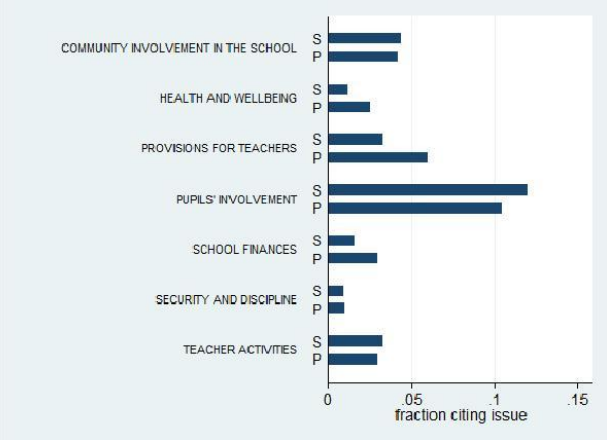
Notes: Figure shows fraction of respondents answering 'yes' to questions concerning various measures of scorecard participation. Results are broken down by standard treatment (S) and participatory treatment (P). Sample is restricted to scorecard committee members.

Balancing the need for broad-based acceptance of the exercise with the possibility that the process will be captured by powerful interests within the school appears to be an important challenge in designing a school-based monitoring instrument. This is reflected in Figures 5(a) and (b). These reflect participants' responses to questions of issues that received too much, or too little, attention in the scorecard. Strikingly, many of the same issues feature heavily on both lists. This suggests that perhaps an important feature of the participatory approach is to provide a forum to balance these interests.

Figure 5: Satisfaction with scorecard content



(a) Issues receiving too little attention



(b) Issues receiving too much attention

6. Conclusions and policy recommendations

This experiment has tested two variants of a 'scorecard' information-for-accountability intervention (Bruns et al. 2011): a standard and a participatory approach, where the latter allowed SMC members to design school scorecards themselves.

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 insignificant. Although the small sample size of our pilot experiment has limited power, in a few cases (such as pupil presence, with controls for baseline characteristics) the differences between the two treatments are statistically significant. Coupled with the remarkably consistent pattern across outcomes, this gives reason to believe that the participatory approach has not only positive impacts, but may also outperform a standard design for such interventions.

There are at least two reasons why this may be the case. 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 provide an opportunity to coordinate "expectations and actions" (Bjorkman and Svensson 2010) of both teachers and parents.

We favour the second interpretation, for two reasons. First, if the participatory approach performed better because of heterogeneity in informational needs, we would expect to see

impacts in particular on intermediate activities-such as school construction-that receive a relatively large share of attention in the participatory scorecard. Instead we find no evidence of increases in expenditure on staff housing or classroom infrastructure under the participatory approach, even though these are the most frequently raised issues in participatory scorecards. Second, under the informational explanation we would not expect the participatory scorecard to outperform the standard scorecard on indicators such as teacher presence that are if any-thing better measured under the standard approach, but in fact we do. Furthermore, findings from the laboratory games added emphasis to the analysis of behaviour of participants. The participatory treatment arm had a higher influence in regard to contribution to a public good across stakeholder types-especially among parents whose ethnicity played an equally important role. Taken together, these findings provide suggestive evidence that the key feature of the participatory approach was that it better engaged the entire 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 little 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

Alderman, Harold, Daniel O. Gilligan and Kim Lehrer, "The impact of food for education programs on school participation in Northern Uganda", forthcoming, *Economic Development and Cultural Change*, 2010.

Andrabi, Tahir, Jishnu Das, and Asim Ijaz Khwaja, "Report cards: The impact of providing school and child test scores on educational markets," Unpublished manuscript, World Bank May 2009.

Andreoni, James, "Why free ride? Strategies and learning in public goods experiments," *Journal of Public Economics*, 1988, 37 (3), 291-304.

Banerjee, Abhijit, 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.

Barr, Abigail and Andrew Zeitlin, "Dictator games in the lab and in nature: External validity tested and investigated in Ugandan primary schools," CSAE Working Paper Series, No. 2010-11 May 2010.

_____ **and** _____, "Conflict of interest as a barrier to local accountability." CSAE Working Paper series, No. 2011-13 August 2011.

Bjorkman, 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.

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.

Deininger, Klaus, "Does cost of schooling affect enrolment by the poor? Universal primary education in Uganda," *Economics of Education Review*, 2003, 22, 291-305.

Duflo, Esther and Rema Hanna, "Monitoring Works: Getting Teachers to Come to School", Mimeo, MIT May 2006.

_____**, Pascaline Dupas, and Michael Kremer**, "Additional resources versus organizational changes in education: Experimental evidence from Kenya," Unpublished, MIT May 2009.

_____**,Rachel Glennerster, and Michael Kremer**, "Using Randomization in Development Economics Research: A Toolkit," Centre for Economic Policy Research, Discussion Paper No. 6059 January 2007.

Fischbacher, Urs and Simon Gachter, "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.

Glewwe, Paul, Michael Kremer, Sylvie Moulin, and Eric Zitzewitz, "Retrospective vs. prospective analyses of school inputs: the case of flip charts in Kenya," *Journal of Development Economics*, June 2004, 74 (1), 251-268.

_____,**Nauman Ilias, and Michael Kremer**, “Teacher incentives," *American Economic Journal: Applied Economics*, July 2010, 2 (3), 205-227.

Guloba, Madina and J C Okuk Nyankori, “Performance of School Management Committees,” Paper commissioned by Improving Institutions for Pro-Poor Growth Research Consortium, project on Management and Motivation in Ugandan Primary Schools June 2010.

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.

_____**,**“Does social capital promote participation in school governance? Evidence from four districts in Uganda," Paper commissioned by Improving Institutions for Pro-Poor Growth Research Consortium, project on Management and Motivation in Ugandan Primary Schools June 2010.

Kremer, Michael, Edward Miguel, and Rebecca Thornton, “Incentives to learn,” *Review of Economics and Statistics*, February 2009, 91 (3), 437-456.

Lassibille, Gerard, Jee-Peng Tan, Cornelia Jesse, and Trang Van Nguyen, “Managing for results in primary education in Madagascar: Evaluating the impact of selected workflow interventions," *World Bank Economic Review*, 2010, 24 (2), 303-329.

Lee, David, “Trimming for Bounds on Treatment Effects with Missing Outcomes,” NBER Technical Working Paper No. 277 June 2002.

Muralidharan, Karthik and Venkatesh Sundararaman, “The impact of diagnostic feedback to teachers on student learning: Experimental evidence from India,” *Economic Journal*, 2010, 120 (5466), F187-F203.

Pandey, Priyanka, Sangeeta Goyal, and Venkatesh Sundararaman, “Community Participation in Public Schools: The Impact of Information Campaigns in Three Indian States,” World Bank Policy Research Working Paper No. 4776 November 2008.

Piper, Benjamin and Medina Korda, “EGRA Plus: Liberia,” Program Evaluation Report draft, RTI International 2010.

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

Appendix A: Scorecard format

Appendix Figure A.1: Participatory design scorecard

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

Appendix Figure A.2: Standard scorecard