Management and motivation in Ugandan primary schools
Report on baseline survey

Economic Policy Research Centre
Uganda

and

Centre for the Study of African Economies
University of Oxford, UK

Final Paper, May, 2010

Contents

1 Background

The PEP-IIG project in Uganda is piloting and evaluating policies to strengthen School Management Committees as a means to improve performance in government primary schools. A randomized experiment design is being used to evaluate the policies. This was in response to the observation that despite high levels of school enrollment following the introduction of Universal Primary Education, completion rates were as low as 22.5 percent in 2006. In 1997, the Government of Uganda introduced Education For All (EFA) at primary level as part of the efforts to universalize primary education and to meet the Millennium Development Goal (MDG) 2 and 3 by 2015. School fees and other additional charges such as contribution to the Parents Teachers Association (PTA) were abolished. The abolition of fees increased school enrolment by over 73 percent nationwide in one year and by 2008, enrolment increased by about 6 percent from 2007. In addition to low completion, other indicators such as school absenteeism are not as expected.

In June 2008, the Improving Institutions for Growth (IIG) project in Uganda conducted a baseline data collection exercise on the school, the school management, pupils, parents and teachers. In February 2009, a partnership was entered with The Netherland Organisation (SNV) Uganda and World Vision (WV) Uganda to undertake policy interventions to strengthen School Management Committees. This baseline report provides highlights of the process and emerging results.

Project objectives

- To identify policies to improve the quality of education in Ugandan government funded primary schools. The national scope of the project, as well as close collaboration with the Ministry of Education and Sports (MoES), will ensure maximum potential for scalability. Further, conducting a comparable intervention in diverse districts will allow us to examine the extent of external validity problems: and
- To mainstream rigorous evaluation of future policies within the office of the Commissioner for Education Planning in Uganda.

2 Methodology

a) Data collection process (sample design & selection)

Four districts were purposefully selected for participation in the baseline and subsequent intervention, with the intention of bringing out some of the challenges of low-performing schools in each of the four regions of Uganda, as well as the availability of implementing partners and other logistical considerations. In each of the resulting four districts, Apac, Iganga, Hoima and Kiboga, data were collected on 25 primary schools. A two-stage sampling procedure was used, first drawing a set of 5 subcounties and then drawing 5 schools from within each of these. Weighting at each stage by the pupil population ensured that selected schools are statistically representative of rural school-going pupils.

b) Field strategy

Data collection was led in each district by a researcher from Economic Policy Research Centre (EPRC), with Centre for the Study of African Economies (CSAE) providing overall support for implementation. Teams of enumerators in each district were supplied by the Uganda Bureau of Statistics (UBoS), which participated jointly in the training of field staff. District education and MoES officials were involved at the initial stages of the data collection process. They were involved in the training on impact evaluation, development of indicators on which data would be collected, actual design of the instruments and field data collection to enable the ownership and usage of the research results.

3 Data structure & school characteristics

In each school, four types of survey instrument were used. These are briefly described below before presenting the descriptive results.

First, a school-level instrument collected basic data on the composition of the school, past performance, inputs such as scholastic materials, and the governance of the school.

Second, in conjunction with officers of the National Assessment of Progress in Education (NAPE), standardized tests of pupil learning in literacy and numeracy were administered to 20 randomly selected pupils in each of primary 3 and primary 6 levels. The NAPE exams are annually administered to monitor performance in primary schools; NAPE is a branch of the Uganda National Examination Board (UNEB) mandated by the government of Uganda to conduct such tests and examinations across the country. Use of NAPE exams provides a reliable measure of educational quality that is well known to the policy community.

Third, individual surveys were administered to a sample of key stakeholders in the school. These were randomly sampled from four populations: the head teacher, the school

management committee, the teachers, and the parents. In addition to a set of basic socio-economic data, these individual questionnaires captured the nature of the relationship of each party to the school. In the case of SMC members, this included date of the most recent meeting held, in which capacity the SMC member is representing on the committee, whether they do have a child in the school among others

Finally, a set of behavioral games were played to measure motivation and strategic interactions among teachers, parents and school management committee members. The descriptive features of school, test, and behavioral game data are presented in the following subsections.

3.1 School questionnaires

The school questionnaire provides information on, among other things, enrollment and performance of pupils in each school and physical and human resources available.

Data on enrollment are presented in Table 1, which displays significant heterogeneity across districts. Class sizes are smallest in Kiboga district (in spite of its poor performance). In all districts, there is significant dropout over time; one measure of this is that class sizes in P7 are approximately one third of those in P1. These patterns are broadly consistent for both male and female pupils.

Table 1. Enrollment by district and primary level

		Apac		Hoima		Iganga		Kiboga	
		mean	sd	mean	sd	mean	sd	mean	sd
P1	male	71.36	(26.43)	111.56	(225.81)	65.92	(24.58)	39.05	(21.77)
	female	72.52	(28.54)	52.47	(43.33)	68.29	(26.07)	34.89	(22.22)
P2	male	60.56	(25.51)	45.88	(31.51)	50.92	(21.89)	25.22	(12.68)
	female	63.20	(26.19)	41.12	(23.38)	52.00	(23.83)	26.78	(16.98)
P3	male	60.40	(25.36)	44.06	(23.03)	50.08	(19.71)	26.28	(15.25)
	female	62.44	(26.83)	41.76	(22.23)	51.63	(21.47)	25.67	(15.42)
P4	male	60.84	(22.49)	40.59	(23.32)	47.88	(19.99)	24.83	(13.67)
	female	63.16	(25.81)	33.24	(18.12)	50.92	(22.36)	26.11	(12.00)
P5	male	57.52	(24.31)	36.18	(16.13)	45.17	(18.96)	20.50	(13.01)
	female	57.28	(29.87)	31.53	(15.82)	42.46	(19.37)	20.56	(14.81)
P6	male	48.08	(23.55)	30.35	(15.96)	33.04	(13.77)	13.65	(10.69)
	female	49.16	(31.99)	27.00	(13.27)	35.04	(18.31)	16.78	(13.82)
P7	male	29.20	(15.48)	17.71	(10.68)	22.38	(14.25)	12.78	(12.20)
	female	22.88	(13.64)	16.00	(10.90)	21.13	(13.26)	13.28	(11.97)

Source: Baseline data, 2008

Table 2 presents results from the Primary Leaving Exam (PLE) in each district. The PLE is the main assessment tool for pupils and determines their progression to secondary school. As shown in Annex B-Figure 1 as well, results at the highest level are extremely rare in all districts. However, at Division 2 level, there is one clear and surprisingly outlier: more than 50 percent of pupils in Hoima district achieve Division 2 results. This finding is borne out in MoES data as well, where Hoima has shown a clear upward performance trend in the last decade. ***The reasons as to how Hoima has so dramatically outperformed neighboring districts such as Kiboga remain an area of research interest.***

Table 2. PLE results by district (%)

	Apac		Hoima		Iganga		Kiboga	
	mean	sd	mean	sd	mean	sd	mean	sd
Division 1	0.01	(0.02)	0.03	(0.09)	0.01	(0.03)	0.01	(0.01)
Division 2	0.29	(0.14)	0.52	(0.18)	0.20	(0.14)	0.26	(0.20)
Division 3	0.24	(0.11)	0.25	(0.11)	0.23	(0.11)	0.25	(0.12)
Division 4	0.20	(0.09)	0.07	(0.05)	0.19	(0.14)	0.18	(0.13)
Division U	0.21	(0.17)	0.09	(0.06)	0.26	(0.16)	0.21	(0.14)
Division X	0.05	(0.05)	0.04	(0.05)	0.10	(0.07)	0.10	(0.07)

Note: not all schools had candidates for the PLE in 2007; for example, several schools do not offer Primary 7.

Source: Baseline data, 2008

Staffing problems vary by district, as shown in Table 3. The average number of teachers per school varies from 8.10 in Kiboga to 10.68 in Apac. However, given the heterogeneity in class sizes there remains significant variation in pupil-teacher ratios across districts (see also Figure 3 for a more complete illustration of the variation in pupil-teacher ratios across districts).

It is also notable that, while districts evidently vary in their ability to fill teacher vacancies, they are also not all equally able to find well qualified teachers. To see this, we consider the fraction of total teachers who have obtained a Grade V qualification (the highest level of teacher certification). We make this comparison by dividing the average number of Grade V teachers per school by the average number of teachers in total. For example, only 12 percent of teachers in Kiboga have a Grade V qualification, whereas 21 percent of teachers in Apac have the same advanced level of qualification. Correspondingly, the use of licensed teachers (LT) to fill vacancies is significantly higher in Kiboga. The observed ability both to achieve higher staffing levels and to fill these positions with qualified staff is somewhat surprising given the relatively remote nature of many of the sampled schools in Apac. Iganga as well presents a very high fraction of Grade V teachers (34 percent), which may go some distance to explaining the relatively strong performance of pupils in that district.

Table 3 . Teacher qualifications, by district

3.2	Apac		Hoima		Iganga		Kiboga	
	mean	sd	mean	sd	mean	sd	mean	Sd
number of teachers	10.68	(2.46)	9.74	(2.16)	9.75	(3.55)	8.10	(2.20)
GT	0.42	(0.64)	0.11	(0.32)	0.28	(0.46)	0.24	(0.54)
grade V	2.23	(1.31)	1.58	(1.50)	3.32	(2.58)	1.00	(0.84)
grade IV	0.00	(0.00)	0.47	(1.43)	0.12	(0.33)	0.00	(0.00)
grade III	7.42	(2.64)	6.47	(2.61)	5.28	(2.15)	5.24	(2.32)
grade III	0.00	(0.00)	0.00	(0.00)	0.04	(0.20)	0.05	(0.22)
grade I	0.00	(0.00)	0.05	(0.23)	0.00	(0.00)	0.00	(0.00)
LT	0.08	(0.27)	0.26	(0.56)	0.00	(0.00)	0.48	(0.81)
pupil-teacher ratio	72.78	(23.01)	53.92	(25.98)	67.76	(20.91)	34.15	(19.82)

Note: rows 2 – 8 present the mean number of teachers with each level of qualification in each school.

Source: Baseline data, 2008

Table 4 summarizes key input ratios by district. There appear to be greater disparities in these physical resources than in the human resources across schools. In particular, pupil-classroom ratios range from sample averages of 129 and 143 in Apac and Iganga respectively (nearly twice the government target of getting this ratio below 80) to 45 in Kiboga. The availability of textbooks follows a similar pattern, with Apac and Iganga significantly behind Hoima and Kiboga. Since regular testing of pupils is not nationally practiced as recommended, and since funding for the exam scripts required to conduct such tests, testing must often be raised outside of the normal school budget. We also report the fraction of schools that have spent any money on the testing of their pupils in the academic year as of the date of the survey. Testing practices vary widely by district, with low levels of testing in Apac an apparent outlier.

Table 4. Input ratios

	Apac		Hoima		Iganga		Kiboga	
input ratio	mean	sd	mean	sd	mean	sd	mean	Sd
pupils per classroom	129.09	(75.27)	63.25	(31.02)	143.28	(90.92)	45.11	(26.17)
pupils per textbook	2.64	(7.46)	1.03	(1.22)	1.25	(1.84)	0.46	(0.31)
any money spent on tests	0.46	(0.51)	0.80	(0.41)	0.80	(0.41)	0.65	(0.49)

Source: Baseline data, 2008

3.3 NAPE exams

In collaboration with the survey teams, officials from UNEB carried out standardized tests in all sample schools. Tests were administered to pupils in primary 3 and primary 6 and in Literacy and Numeracy.

The examination instruments used were those of the 2006 National Assessment of Progress in Education (NAPE). This is an annual testing exercise undertaken on a sample of Ugandan primary schools by the NAPE department on behalf of UNEB (the NAPE sample size was 405 schools in 2006). In our procedure as in that of the NAPE assessment, 20 pupils were randomly selected from each of the P3 and P6 classes for participation in the tests. The grading of the exams was done by NAPE using the UNEB guidelines in assessment.

Table 5. NAPE exam results by district and year

	P6		P3	
District	numeracy	literacy	numeracy	literacy
Iganga	18.54 (10.94)	13.97 (9.50)	11.70 (6.67)	10.20 (6.68)
Apac	25.27 (11.61)	16.66 (10.73)	16.79 (9.92)	13.86 (7.20)
Hoima	28.42 (13.59)	23.98 (13.43)	23.14 (9.43)	18.29 (7.63)
Kiboga	31.67 (11.92)	23.61 (10.91)	21.31 (8.14)	17.95 (7.23)
Total	25.52 (12.94)	19.17 (11.98)	18.17 (9.72)	15.00 (7.92)

Note: P3 exam results range from 0-50; P6 results range from 0-100. Figures in the parenthesis are standard deviations.

Source: Baseline data, 2008

Test results are presented in Table 5. Levels of achievement are low across all study schools, with problems particularly severe - somewhat surprisingly – in Iganga. The external validity of these test results will be taken up in 4.1 below, where it is shown that they are strongly predictive of PLE results for each school per district.

3.4 Game data

Evidence suggests that teacher motivation matters for school performance. In a low accountability environment, where teachers are rarely punished for absenteeism,[1] forms of intrinsic motivation may be just as important as contractual incentives and career concerns in motivating teachers to perform. Economic theory has suggested that if intrinsic motivation is important, the use of high-powered incentive contracts may actually worsen performance – a "crowding out" effect. Thus, it is important to understand the role of intrinsic motivation.

While motivation is difficult to measure with standard survey instruments, an alternative approach comes from the laboratory experiments of behavioral economics. *Behavioral games* provide a means to measuring motivation and strategic interactions among individuals. Since such games have been widely played, they have the advantage of external validity: they measure well established aspects of individuals preferences, aspects that are known to be correlated with real-world characteristics.

We played two types of standard behavioral games with the individuals interviewed.
The first of these is a Dictator Game (DG). The DG is played between pairs of individuals, who are randomly matched and do now know with whom they are playing. In each pair, one plays the role of 'dictator' and the other is passive. The dictator is given a finite sum of money to divide among the two however she likes (in this case the sum was UShs 5,000 in each round). She simply makes her division, keeping as much or as little as she likes, and at the end of the experimental session the individuals keep whatever shares she allocated them. The DG therefore provides a measure of the strength of the dictator's regard for the allocation to the recipient. In contrast to the predictions of models assuming self-interested and rational agents, it has been found that individuals share a significant portion of the endowment with recipients across a wide variety of cultural contexts (Henrich et al. 2006).

The second game, a Third Party Punishment Game (3PPG), introduces a twist onto the DG. The game is now played by three randomly matched, mutually anonymous individuals. To begin, a dictator is given a fixed amount of money (again, UShs 5,000) to divide between herself and a passive recipient, just as in the DG. However, a third party is then given a chance to express disapproval of the dictator's allocation. This third party has an initial endowment of UShs 2,500 in this game. If the third party is unhappy with the dictator's allocation, they can choose to spend a fraction (UShs 500) of their own endowment in order to have the dictator fined. If the dictator is fined, she loses UShs 1,500 from whatever she chose to keep for herself. The 3PPG

[1] For example, Chaudhury et al. (Chaudhury et al. 2006) find that 27 percent of teachers are absent from Ugandan primary schools on any given day.

therefore provides a measure of the third party's willingness to enforce their notion of 'fair' allocations between the dictator and recipient. To protect anonymity and maximize the information contained in their responses, third parties are asked about their willingness to fine *before* being told the specific decision of the dictator with whom they were matched. Third parties express their willingness to fine by deciding, for each possible allocation that could be made by the dictator, whether they would like to fine.

There was one distinction between the way that these games were played among survey participants and the way that they are typically played in a research laboratory setting. Rather than being randomly assigned to roles (dictator, recipient, third party), participants were assigned to roles on the basis of their relationship to the school. Teachers were cast as dictators; parents were cast as third parties; and head teachers and SMC members were cast as third parties. Consequently the game provides insight into an aspect of the extent to which teachers 'care' about the wellbeing of local community members; this is hypothesized to play a role in their intrinsic motivation in the classroom. The game also sheds light on the willingness of head teachers and SMC members to punish what they perceive to be unfair behavior by teachers.

Teachers' allocations to parents are presented in Table 6. The first row expresses the percentage of the total allocation (UShs 5,000) allocated to parents on average in each district (standard deviations in parentheses). Teachers allocate an average of approximately 40 percent of their endowment to parents, a feature which is broadly consistent across districts. As Figure 4 shows, the most common allocation is a gift of 50 percent to the corresponding parent; this is the choice made by approximately 31 percent of teachers. The vast majorities (82 percent) of teachers give 50 percent or less of the endowments to the corresponding parent, and 6 percent of teachers give nothing at all.

Table 6 . Dictator and Third Party Punishment Games (%)

	Apac	Hoima	Iganga	Kiboga
teacher allocation to parent, DG	0.42	0.42	0.35	0.39
	(0.21)	(0.22)	(0.20)	(0.16)
teacher allocation to parent, 3PPG	0.42	0.37	0.36	0.37
	(0.22)	(0.23)	(0.25)	(0.21)
minimum unfined allocation, 3PPG	0.43	0.35	0.57	0.29
	(0.18)	(0.16)	(0.33)	(0.20)

Note: Decisions are presented as percentages of the total endowment under the control of the dictator. Standard deviations in parentheses.

Source: Baseline data, 2008

Teachers did not significantly alter their allocations to parents when faced with the possibility of being fined by a third party. This suggests that the threat of the fine did not generally induce a strong increase in gifts given by teachers.

The third row of Table 6 shows the minimum unfined allocation by the third parties (head teachers and SMC). This shows the smallest share that the teacher could have given to her corresponding parent without the SMC member deciding to fine them for keeping so much to themselves. As shown in the table and in Source: Baseline data, 2008

Figure 5, SMC members on average typically allow an offer of as little as 41 percent to go unpunished. About 1/3 of SMC members allow offers of less than 40 percent to go unpunished. Interestingly, the decisions of teachers and SMC members – which are made independently of one another – are negatively correlated. Figure 6 shows the negative correlation between teacher allocations and minimum unfined allocations; this relationship is statistically significant at the 10 percent level. Schools where management is more likely to punish even relatively generous offers by teachers are schools where teachers tend to offer less. While great caution is required in interpreting such simple correlations, this seems consistent with the view that there is a breakdown in social capital and morale in poor performing schools – especially in light of the evidence of Section 4.

4 Correlates of education quality

In this section we examine the correlates of education quality in PLE schools. We do so in two steps. First, we show that our measure of education quality – based on the NAPE exams – is correlated with PLE results and other measures of education quality. Second, having established the external validity of our performance measure, we examine the correlates of school performance, with attention to three broad features of the school: physical inputs, teachers, and management and motivation.

The results presented here must be interpreted with great caution. This is particularly the case because, as is well understood, observed correlations between educational inputs and school performance can not be interpreted as causal in general. For example, Glewwe et al. (Glewwe et al. 2004) have shown that the educational returns to the use of flipcharts in Kenyan primary classrooms as estimated based on observational data are significantly different than estimates based on randomized controlled trials (overestimated, in that case, though there is no reason to believe the bias will be in the same direction in other cases). In fact even the observational estimates that Glewwe et al. consider should be more robust than the correlations that we can

present here, because they are able to use panel data to apply a difference-in-differences approach.

4.1 How do the NAPE tests correlate with PLE performance?

The NAPE exams administered are strong predictors of PLE performance, as shown in Table 7. Each of the table's two panels gives the results of four separate regressions. Within a panel, all regressions regress a single dependent variable on four measures of school quality from the NAPE exams: primary 3 literacy results, primary 3 numeracy results, primary 6 literacy results, and primary 6 numeracy results. These exam scores are standardized (so that all have mean zero and variance of one) by exam type and averaged within schools. Consequently the estimated coefficients are comparable across tests within a given panel.

Panel A regresses the percentage of pupils receiving Division I or II scores on the NAPE exam results. The results can be interpreted as marginal effects (with care about causality): for example, schools where NAPE exams are one standard deviation higher are schools with 11 percent more pupils achieving results of Division I or II. Given that schools on average have only 32 percent of their pupils in these two upper divisions, these measures of school quality appear to be very strong predictors of school success.

Table 7 . Correlation between NAPE tests and PLE exam results

Panel A - Dependent variable: Percent pupils achieve Division I or II scores			
P3 Literacy	P3 Numeracy	P6 Literacy	P6 Numeracy
0.11***	0.11***	0.20***	0.14***
(0.03)	(0.03)	(0.04)	(0.04)
Panel B - Dependent variable: percent pupils achieve Division X or U scores			
P3 Literacy	P3 Numeracy	P6 Literacy	P6 Numeracy
-0.08***	-0.08***	-0.14***	-0.11***
(0.03)	(0.02)	(0.04)	(0.03)

Source: Baseline data, 2008

By contrast, the dependent variable in Panel B is a measure of *poor* performance: it is the percentage of pupils receiving incomplete or failing marks on their PLE. Again, the NAPE exams are strongly correlated with this measure of school quality. The coefficients imply, for example, that schools that are one standard deviation lower in P3 literacy scores have an average of 8 percent fewer pupils in Division X or U (compare with 27 percent of pupils in these bottom divisions on average across all schools). Taken together with the results of the upper panel, this

suggests that the exams are sensitive measures of performance at both the upper and lower end of the distribution.

4.2 Production of education quality

Having given evidence of the external validity of our measure of education quality, we now turn to examine its correlates among the inputs and managerial measures of the school. Table 8 estimates an education production function, with factor inputs, teacher quality measures, and management/motivation on the right hand side; this is done separately for each of the four measures of test scores.

Table 8 . Education production function

	(1) P3 Lit	(2) P3 Num	(3) P5 Lit	(4) P6 Num
ln enrollment	-0.27	-0.20	-0.05	0.09
	(0.16)	(0.18)	(0.13)	(0.14)
ln teachers	0.04	0.06	-0.10	-0.34
	(0.35)	(0.39)	(0.29)	(0.31)
pct teachers with GT qualification	1.16	0.04	-1.43	-0.04
	(1.55)	(1.73)	(1.31)	(1.36)
ln classrooms	0.26	0.21	0.281*	0.18
	(0.17)	(0.20)	(0.15)	(0.15)
ln texts	-0.132*	-0.151*	-0.04	-0.02
	(0.08)	(0.09)	(0.07)	(0.07)
teacher allocation in DG	-0.14	0.26	-0.02	-0.09
	(0.56)	(0.63)	(0.47)	(0.49)
SMC minimum unfined allocation in 3PPG	-0.77**	-0.86**	-0.64**	-0.90***
	(0.36)	(0.40)	(0.31)	(0.31)
N	80	80	78	80

Note: standard errors in parentheses. *, **, *** denote significance at 10, 5, 1 percent level.
Source: Baseline data, 2008

As is found elsewhere in the literature, such a cross-sectional approach does not yield strong correlations between inputs – physical and human resources – and learning outcomes. The relationship between physical infrastructure (classrooms) and pupil achievement is significant in one case; perhaps surprisingly, textbooks enter with the wrong sign in two of the regressions.

What is striking, however, is the correlation between the play in the behavioral games and the performance of pupils. The statistically significant and robust relationship between SMC decisions in the third party punishment game and the performance of pupils implies, for example, that an increase from 0 to 1 in the minimum unfined allocation to parents is

correlated with a 0.77 standard deviation decrease in pupil performance. Recall that the higher the minimum unfined allocation, the less willing is the management committee to punish self-interested behavior on the part of teachers. This suggests that the combined effect of management holding views of fairness that serve to protect parents and their being willing to stand up to enforce these is potentially an important factor in the efficiency with which inputs are put to productive use in the school.

5 Covariate balance by treatment arms

While successful randomization ensures that, asymptotically, both observed and unobserved characteristics are balanced across the control group and the two treatment groups, it is possible in small samples for there to be differences between the two. This is the case even though the randomization was "successful" in a practical sense: there was no form of interference that intervened between the initial, random assignment and the allocation of schools to treatment arms.

Consequently, it may be useful to check that observed outcomes are balanced across treatment groups. If there are (small-sample) differences across groups, then in addition to reporting results of the evaluation based on the randomized assignment alone (without controls), it may be a valuable robustness check to include regressions that include controls for key characteristics. Chief among these are characteristics that are both important correlates of the outcomes under study and that are unbalanced across treatment arms at the time of the baseline.

This applies *a fortiori* to the baseline values of the outcome measures under study. However, it is our intention to employ a (school-level) *difference-in-differences* estimator to improve precision in the primary estimates of program impacts, even in the absence of any evidence of covariate imbalance at baseline.

As Imbens and Wooldridge (Imbens & Wooldridge 2009) point out, the presence of imbalance in observed characteristics does not imply the presence of imbalance in unobserved characteristics, or vice-versa. However, to the extent that regression will be used to control for differences in observed characteristics across treatment arms, Imbens and Wooldridge argue that as covariates become increasingly unbalanced, estimates become more sensitive to the specification of the regression used.

With the caveats above, we check for covariate balance in the following characteristics: Our standardized (NAPE) tests of literacy and numeracy conducted in all schools; failure and first-

class rates in the Primary Leaving Exam (PLE), teacher absenteeism (defined as mean days absent in our random sample of teachers); pupil-teacher, pupil-classroom, and pupil-textbook ratios, and teachers' allocations in the dictator game. Because we are interested not only in making comparisons between each of the two treatment arms and the control group, we present estimated differences (and associated standard errors) for three, pairwise comparisons: between each treatment group and the control, and between the two treatment groups.

Broadly, the evidence in Table 9 suggests that the randomization has achieved balance in observable characteristics of interest. We find statistically significant differences in only one of our pairwise comparisons: teachers in the first treatment arm exhibit a higher absence rate than teachers in the control group.

It should be noted, however, that our relatively small sample (at the school level) means that these tests are of limited power. This will, of course, be a challenge for the detection of mean program impacts as well, a fact which highlights the importance of (a) controlling for baseline outcomes to improve precision, and (b) examining heterogeneous response at the individual level. With regard to the latter, we are particularly interested in examining whether teachers who exhibit high 'motivation' at baseline (measured by their absenteeism rates or their DG offers) are less responsive to the intervention: a motivational crowding out hypothesis.

Table 9 Covariate balance across treatment arms

	T0	T1	T2	T1-T0		T2-T0	T2-T1
	(1)	(2)	(3)	(4)		(5)	(6)
P3 literacy (NAPE)	15.62	14.91	14.58	-0.71		-1.05	-0.34
	(5.67)	(4.80)	(4.81)	(1.25)		(1.25)	(1.24)
P3 numeracy (NAPE)	18.27	18.19	18.39	-0.08		0.12	0.20
	(6.96)	(7.09)	(7.22)	(1.71)		(1.71)	(1.85)
P6 literacy (NAPE)	20.12	19.23	18.41	-0.89		-1.71	-0.82
	(6.19)	(7.01)	(6.44)	(1.60)		(1.67)	(1.84)
P6 numeracy (NAPE)	26.16	26.03	24.75	-0.13		-1.41	-1.27
	(6.35)	(8.77)	(7.10)	(1.81)		(1.84)	(2.14)
pct PLE div. 1	0.01	0.02	0.01	0.01		0.00	-0.01
	(0.02)	(0.07)	(0.02)	(0.01)		(0.01)	(0.01)
pct fail PLE	0.30	0.25	0.26	-0.05		-0.04	0.01
	(0.17)	(0.17)	(0.17)	(0.05)		(0.05)	(0.05)
Teacher absences	2.54	3.63	3.21	1.09	**	0.66	-0.43
	(1.11)	(2.38)	(2.06)	(0.45)		(0.45)	(0.57)
DG offer (percent of stake)	0.41	0.41	0.39	0.00		-0.02	-0.02
	(0.12)	(0.14)	(0.12)	(0.03)		(0.03)	(0.03)
Pupil-teacher ratio	56.76	65.71	63.40	8.95		6.64	-2.31
	(24.97)	(25.40)	(25.60)	(6.27)		(6.40)	(6.82)
Pupil-classroom ratio	99.80	116.36	89.44	16.56		-10.36	-26.93
	(72.25)	(94.09)	(52.52)	(18.80)		(19.21)	(20.96)
Pupil-textbook ratio	1.88	1.80	0.83	-0.08		-1.05	-0.97
	(6.13)	(2.97)	(0.62)	(1.05)		(1.10)	(0.60)

Notes: Columns (1), (2), and (3) present means and standard deviations by treatment group. T0 denotes control group; T1 denotes "community-defined" scorecard; T2 denotes "shared format" scorecard. Columns (4), (5), and (6) present differences (and associated standard errors) for comparisons across treatment groups. In these columns, *,**,*** denotes significance at the 10, 5, and 1% levels, respectively.

6 Conclusions and steps forward

Results from the baseline survey are consistent with the view that management plays a central role in determining the success or failure of Ugandan primary schools in producing quality education. This provides support to the analytical approach of the project as a whole, which is implementing a randomized controlled trial to evaluate the impact of an intervention that strengthens the functioning of the School Management Committee in sample schools.

The intervention under study focuses on the use of a School Management Committee Scorecard. This is a monitoring tool designed to provide a focal point for the activities of the SMC. The teachers, parents and other members of the school management committee will complete the scorecard. Through the scorecard, each group will assess school performance on human and physical resources. A consensus forum will decide on one school scorecard. The school scorecard will be a basis for action at the school level. It will also be communicated to the district education office. The district education office will use the scorecard to conduct targeted support supervision visits, thereby using the available merger resources efficiently. The scorecard process will be conducted at least once a term and repeated each term for at least one year. This increased flow of appropriate information and follow up action is expected to improve the management of schools. World Vision and SNV will coordinate the implementation of the intervention. Economic Policy Research Centre and the Centre for Studies of African Economies will conduct the evaluation. In September 2009, the intervention will start in all the four districts: Apac, Iganga, Hoima and Kiboga.

As part of strengthening research capacity in Uganda, IIG Uganda will issue a call for proposals to researchers in Makerere University and other leading academic institutions to collaborate on using project baseline data to advance scientific work.

In July, IIG Uganda will continue in its capacity strengthening in Impact Evaluation with a 3-day training in impact evaluation on improving institutions for medicine delivery in Uganda. IIG in Uganda will also conduct a 3-day training in panel data analysis in Kampala, Uganda in September to coincide with the wider IIG meeting.

IIG Uganda will continue its work on engaging with the policy process in Uganda. Planned activities include, among others, the presentation of baseline results and intervention and analytical design at a workshop on the role of School Management Committees sponsored by implementing partner SNV.

Annex A. References

Chaudhury, N., Hammer, J., Kremer, M., Muralidharan, K. & Rogers, F. H. (2006), 'Missing in action: Teacher and health worker absence in developing countries', *Journal of Economic Perspectives* **20**(1), 91–116.

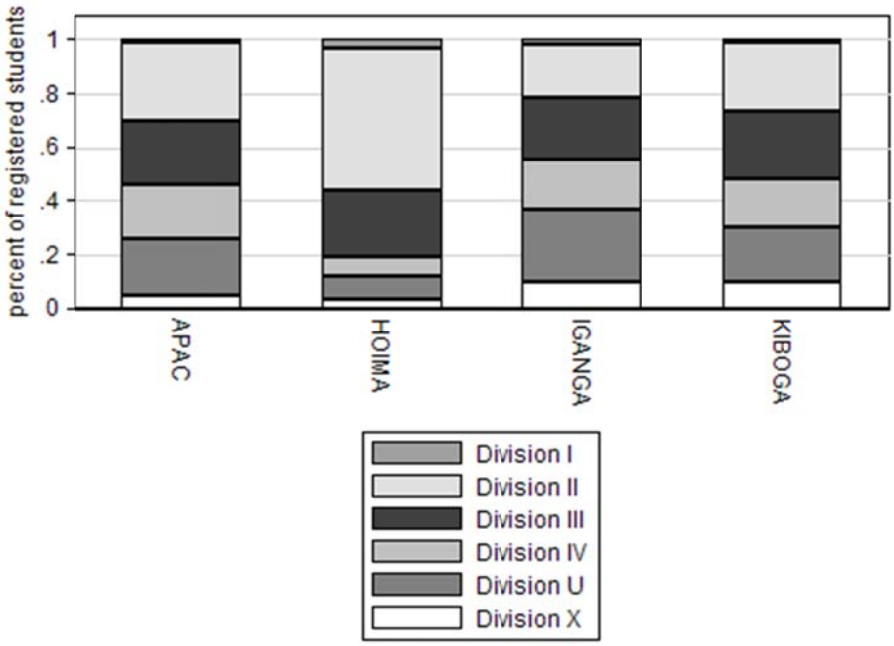
Glewwe, P., Kremer, M., Moulin, S. & Zitzewitz, E. (2004), 'Retrospective vs. prospective analyses of school inputs: the case of flip charts in kenya', *Journal of Development Economics* **74**(1), 251–268.

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., Cardenas, J. C., Gurven, M., Gwako, E., Henrich, N., Lesorogol, C., Marlowe, F., Tracer, D. & Ziker, J. (2006), 'Costly punishment across human societies', *Science* **23**, 1767–1770.

Imbens, G. W. & Wooldridge, J. M. (2009), 'Recent developments in the econometrics of program evaluation', *Journal of Economic Literature* **47**(1), 5–86.

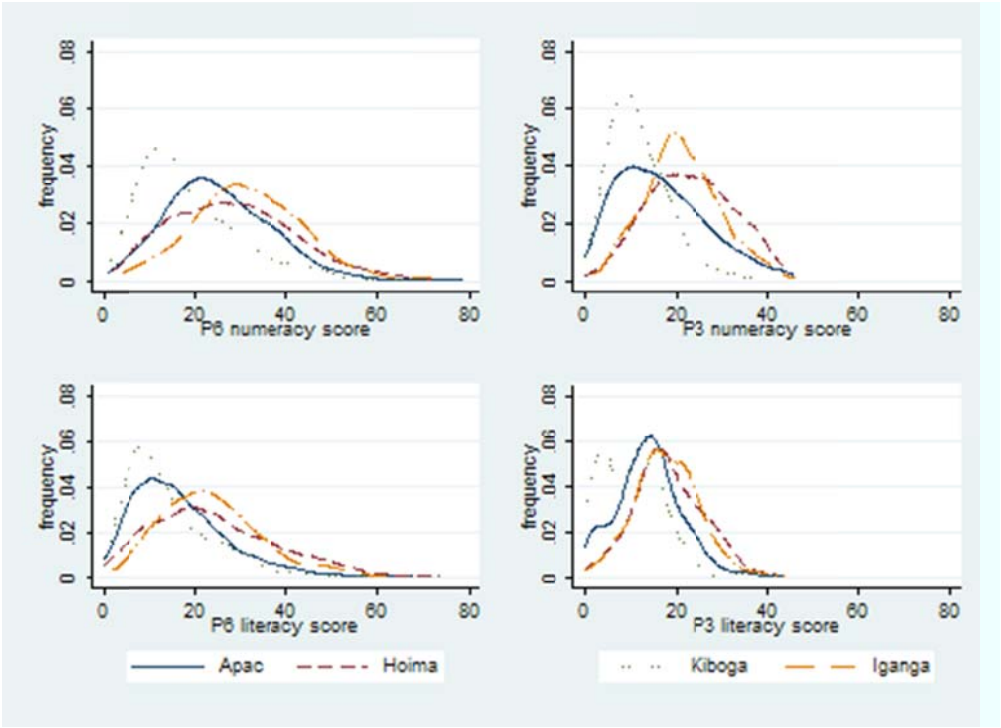
Annex B. Figures

Figure 1. Primary Leaving Exam results by district



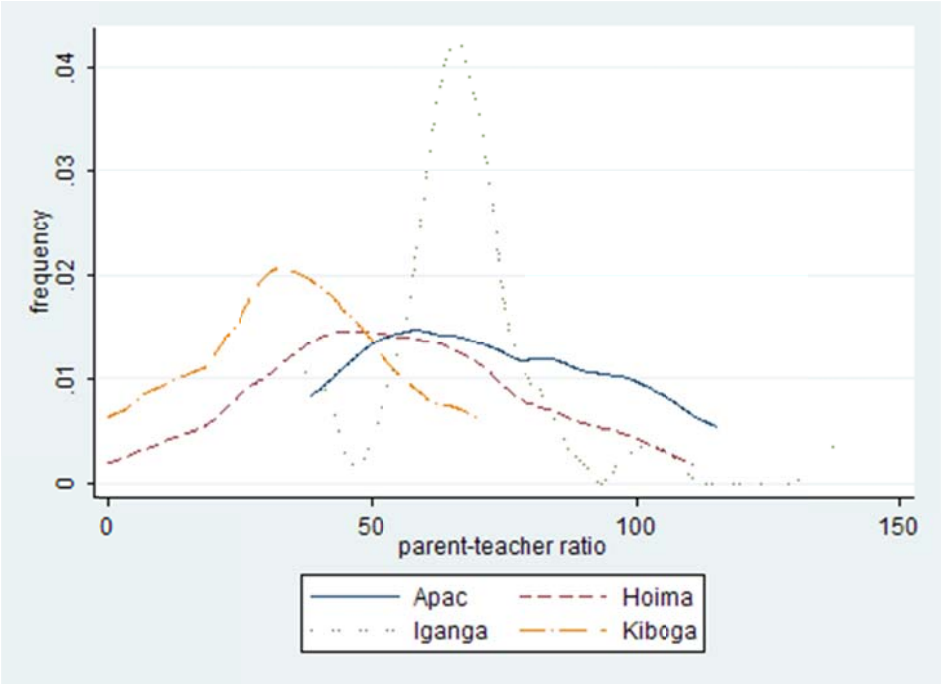
Source: Baseline data, 2008

Figure 2. NAPE exam results by district and year



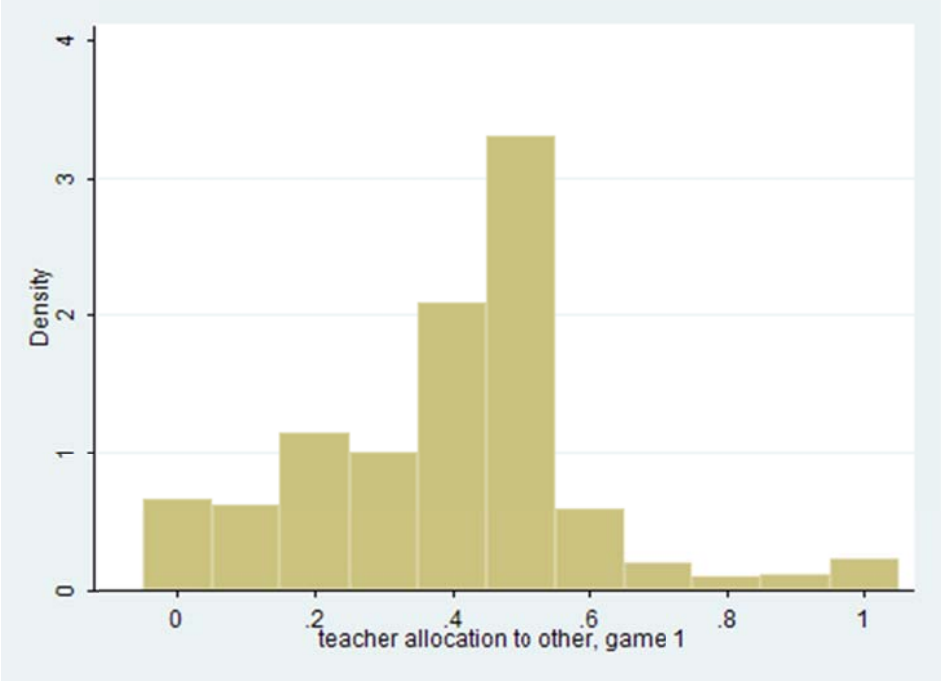
Source: Baseline data, 2008

Figure 3. Pupil-teacher ratio, by district



Source: Baseline data, 2008

Figure 4 . Teacher allocations to parents in Dictator Game



Source: Baseline data, 2008

Figure 5 . Minimum unfined offers in the third party punishment game

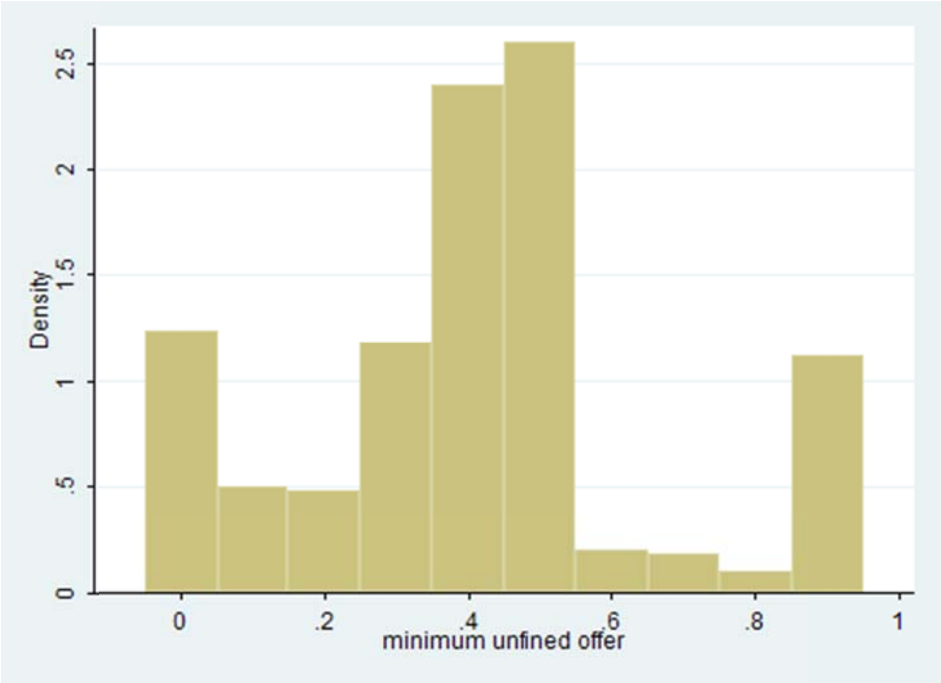
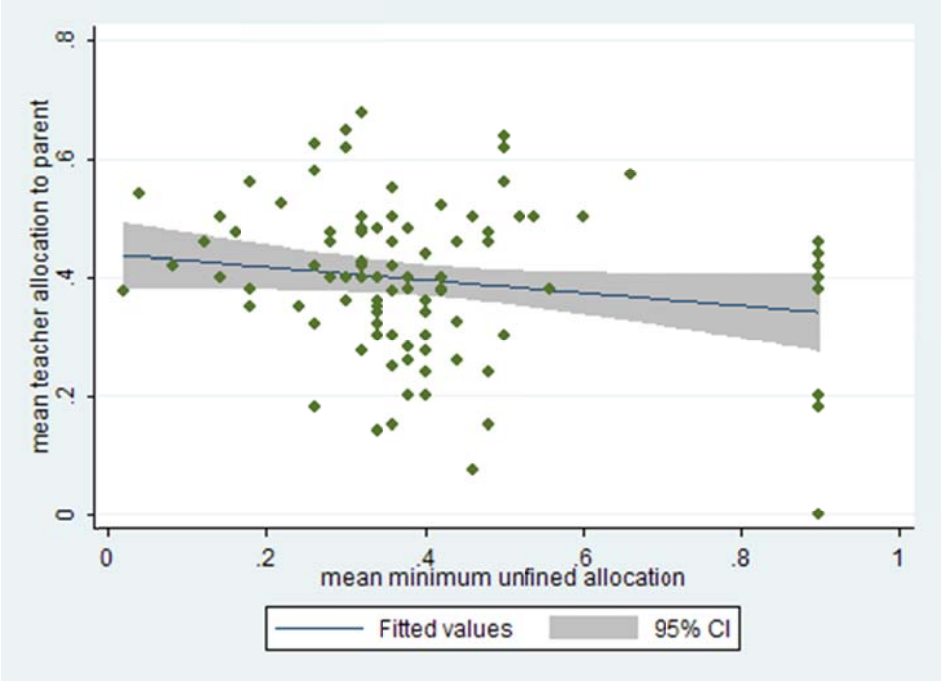


Figure 6. Correlation between SMC fining decisions and teacher offers



Source: Baseline data, 2008