

Community Training, Information, and Demand for Accountability

Pre-Analysis Plan

Nathan Fiala* and Pia Raffler†

Registered: December 14, 2018.

Abstract

In holding their governments to account, citizens face two main problems: first, they often lack information on the extent to which substandard services are caused by the behavior of lower-level officials, rather than factors beyond their control, such as lack of resources or malfeasance at higher levels of government. Second, collective action problems may prevent them from monitoring and holding government officials directly accountable in between elections. These two problems may reinforce each other, resulting in an uninformed and disempowered citizenry. We designed a field experiment that allows us to test the effect of alleviating each of the two constraints mentioned above, or alleviating both of them together, in the context of a large anti-corruption program in Uganda. We focus on demand for accountability, political attitudes, and political behavior as dependent variables.

*Department of Agricultural and Resource Economics, University of Connecticut, nathan.fiala@uconn.edu

†Department of Government, Harvard University, praffler@gov.harvard.edu. The preanalysis plan was registered before analyzing outcome data.

Contents

1	Introduction	3
2	Description of Treatments	4
2.1	Treatment 1: Training of social accountability groups	4
2.2	Treatment 2: Scorecard dissemination	4
3	Experimental Design	6
3.1	Timing and randomization	6
4	Data	7
4.1	Physical project assessment	7
4.2	Endline survey	7
5	Hypotheses	7
5.1	Training	7
5.2	Scorecard	8
6	Key Variables and Measurement	9
6.1	Dependent Variables	9
6.2	Group A: Evaluation of government performance and voting behavior	9
6.3	Group B: Sense of efficacy and collective action	10
6.4	Independent Variables	12
6.4.1	Heterogeneous treatment effects	12
6.4.2	Treatment Variables	12
6.4.3	Covariates	13
6.5	Attrition, Outliers, and Missing Values	13
6.5.1	Attrition	13
6.5.2	Outliers	13
6.5.3	Missing covariate values	14
6.5.4	Missing dependent variables	14
7	Analysis	14
7.1	Main Analysis	14
7.2	Correction for Multiple Hypotheses Testing	15
8	Power Analysis	15
9	Timeline	15
10	Partnership and Human Subjects	15
A	Overview Page of a Sample Scorecard	17

1 Introduction

Poor service delivery by lower-level government officials remains one of the core impediments to development, despite the introduction of multi-party elections in much of the developing world. In holding their governments to account, citizens face two problems: first, they often lack information on the extent to which substandard services are caused by the behavior of lower-level officials on the one hand, and the extent to which they are due to factors beyond their control, such as lack of resources or malfeasance at higher levels of government, on the other. Second, collective action problems may prevent them from monitoring and holding government officials directly accountable in between elections. These two problems may reinforce each other, resulting in an uninformed and disempowered citizenry.

We conduct a field experiment to test the effect of alleviating information and collective action constraints, or alleviating both of them together, on communities' demand for accountability in the context of a large anti-corruption intervention embedded in a community driven development (CDD) program in Uganda. The CDD program gave grants to communities to purchase livestock. We focus on demand for accountability, evaluation of government performance, and political behavior as dependent variables. The first component of the intervention aimed at alleviating collective action problems by having beneficiaries select a small group of representatives, called a social action committee, and giving them the mandate to monitor service delivery. Committee members were trained to monitor the implementation of their local development projects and to address any issues with project implementers and government officials. The second component sought to address the information problem by providing citizens with a project scorecard, including a benchmark on how their local government performed relative to others in the same district. The idea is that the benchmark will provide citizens with a signal about the performance of the actors responsible for implementing their project. Scorecards were based on data from physical audits of all projects to construct simple indicators of the quality of project implementation, as compared to other projects in the same district.

We test the effect of these two interventions in a factorial design in 574 communities receiving livestock through the CDD program. Our interventions were implemented prior to the 2016 elections. Outcomes of interest fall into five categories: evaluation of government performance, sense of efficacy, collective action potential, voting behavior, and taking action aimed at improving services. Endline data is drawn from surveys with 3,851 households.

2 Description of Treatments

2.1 Treatment 1: Training of social accountability groups

We worked with the government of Uganda to embed a randomized control trial into an anti-corruption component of a community driven development (CDD) project. Communities were trained on the details of project implementation and how to identify and prevent cases of corruption and mismanagement. The training was implemented by civil society organizations (CSOs) across northern Uganda. The CSOs sent representatives to communities to train selected community members on social accountability and community monitoring of the CDD projects. The program also supported follow-up visits by CSO representatives to provide on-going training and advise the communities on how to monitor the implementation of their projects.

When the CSO trainers first entered the community, they organized community assemblies. The assemblies discussed the principles of social accountability and community monitoring. As part of this mobilization phase, the community elected representatives to expand an existing social accountability committee. The existing committee was generally considered to be untrained and poorly prepared to monitor issues in the project. Members of the expanded committee made a public pledge to participate in the training program, to undertake monitoring of the project on behalf of the community, and to report back to the community. The training was open to all community members.

The training provided background on social accountability and the CDD program and taught participants how to monitor their projects. The training also provided hands-on skills in writing reports, providing feedback to the community, generating a community action plan, and applying the new monitoring skills to other projects in the community. Special emphasis was put on encouraging communities to reach out and make complaints to local and central government.

2.2 Treatment 2: Scorecard dissemination

Scorecard construction To construct the community scorecard, the research team conducted an assessment of the quality of projects in communities approximately six months after the mean completion date of these projects, in December 2015 and January 2016. The team collected information on both the quality and quantity of the projects through physical observations. This information was then used to construct a score for each community project. In particular, we construct scores for the following four dimensions.

1. Health of animals upon arrival: Health as assessed by beneficiaries by answering a set of specific questions on animal health¹ (50%), reported animal survival rates (50%).

¹These questions were designed with the help of a local veterinarian. Indicators include signs of illness, abnormal

2. Animal productivity: ability to fulfill the intended purpose (give milk, pull a plough, breed) (50%), current health (50%). Score of zero for dead or missing animals.
3. Assistance from the district veterinary officer: Indicators for the six roles that DVO were supposed to complete for each sub-project according to the project documents. These were: 1) follow-up after inspection, 2) animal treatment/prophylaxis 3) animal ear tagging 4) training sub-project committees 5) animal selection 6) animal inspection. The first three roles were asked to survey respondents and we assign a score equal to the fraction of respondents that said the DVO provided that service (e.g. 0.6 if 3 of 5 respondents said the DVO ear tagged their animals). The last three roles were asked during the procurement tool to the sub-project committee.
4. Value for money

The overall score is the weighted average of the above four scores, with the first one, the health of animals upon arrival, weighted double. The construction of these indicators was reviewed by technical government officers. A sample scorecard is included in Appendix A.

Scorecard dissemination Each scorecard meeting was organized by a team of three field officers: one community facilitator, who hosted the meeting, one enumerator, who completed the scorecard survey throughout the meeting, and one mobilizer, who visited the community in advance of the meeting to explain the agenda, schedule the meeting and mobilize community members. Each team held two scorecard dissemination meetings per day.

During the meeting to disseminate scorecards, treatment communities were presented with written and oral information on the ranking of the performance of their project relative to others in the districts; as well as summary information on the health of animals, animal productivity, assistance from the district veterinary officer, who was supposed to assist communities with their animals but was rarely present, and a constructed value-for-money score, which was calculated by multiplying the number of animals received by the productivity score of all the animals, divided by the total amount of money received for the project. All indicators were designed in collaboration with technical officers from the Ugandan government.

Throughout the meeting, communities were invited to discuss the results. This discussion was supported by the community facilitator and included opening remarks from the community leaders and a speech introducing the goals of the meeting. The scorecard results were then announced, with each component of the score fully explained. The meeting ended with a discussion about how communities could use the results of the score to improve service delivery and accountability in the discharges, skin conditions, parasites, temperament and body score.

community. Some of the actions suggested by communities during these meetings included: (1) voicing concerns to the subcounty and district leadership; (2) participating actively in community projects; (3) voting for local politicians whom they believe can best help the community develop; (4) selecting the best possible project leadership and monitoring them closely; and (5) working together as a community to resolve issues whenever they can.

The facilitator brought to each community five copies each of the scorecard in English and the local language, which were left with the community, as well as a number line to graphically show the ranking of the community project relative to others in the district, and sodas and the soap to thank participants for their time (a common practice). Once the facilitators left they did not return to the community.

3 Experimental Design

The factorial research design is summarized in Table 1.

	Community training	No community training
Information	A	B
No information	C	D

Table 1: Factorial Design

3.1 Timing and randomization

The timeline and randomization for this study is as follows. In November 2013, the research team was provided with the list of the projects to be funded. Communities were randomly assigned into social accountability training treatment or control in January 2014, and the program and social accountability trainings began in June 2014. Randomization was implemented in Stata and not stratified. 80% of the funds were distributed in December 2014, with the other 20% funded in the six months preceding this date. The project assessment was conducted from December 2015 to early February 2016. Construction of the scorecard and randomization of communities for the scorecard intervention was done in February, again in Stata. Randomization was stratified on assignment to the social accountability training. The distribution of scorecards took place in February and March 2016. Six months after the assessment, in June and July 2016, the final household survey was conducted. The survey was conducted on a rolling basis to ensure that communities were visited six months after the project assessment.

4 Data

4.1 Physical project assessment

The data used in the construction of the scorecard stems from a physical project assessment, conducted between December 2015 and February 2016. We randomly sampled four beneficiaries from each community. Trained enumerators assessed the animals and, based on their observations, completed a questionnaire developed with a veterinarian. They also conducted a survey with the animals' owners, as well as a separate survey with the members of the project leadership committees. Main outcomes of interest for the assessment are the age of the animal, health, breed, and productivity.

4.2 Endline survey

The endline survey was conducted in June and July 2016. Eight people per community were surveyed, including the two members of the project's executive committee responsible for procurement and project management, respectively, two members of the original community social accountability committee, and two regular beneficiaries². The identity of the remaining two respondents was contingent on treatment assignment. In communities assigned to receive the social accountability training, we interviewed two members from the expanded community accountability committee.³ In communities assigned to receive no social accountability training, we instead interviewed two additional regular beneficiaries. Data from the household survey includes information on assets, including animals and household durables, complaints to local government, evaluation of the performance of different levels of government and the project leadership, sense of efficacy, perceived collective action potential, voting behavior, and the individuals' level of trust in local leaders.

5 Hypotheses

5.1 Training

We hypothesize that being assigned to the training treatment will result in:

- An increased sense of efficacy
- Greater collective action potential

²Regular beneficiaries are members of the group who were supposed to receive livestock, but are not members of any of the project committees.

³To be excluded from the main analyses.

- A greater probability to take action aimed at improving project implementation

Further, as secondary hypotheses, we expect that, conditional on project performance being below the median according to the physical audits, assignment to the training will lead to:

- Decreased trust in and lower perceived performance of low and high level leaders
- A greater probability of voting against the incumbent (party)
- Greater openness towards the opposition

5.2 Scorecard

The scorecard intervention has the potential to affect citizens through two channels: information on a project's performance relative to others in the district and coordination.

First, the scorecard provides information about a project's performance *relative to others* in the district. Given the many different actors involved in service delivery, citizens often have a hard time attributing responsibility correctly. We hypothesize that the scorecard helps citizens understand which level of government to blame for any shortcomings in project implementation. Assuming that constraints are similar within a district, receiving information that one's community project performs below the district median suggests that lower level officials are not performing as well as they could. If, on the other hand, a project performs above the district median, it suggests that lower level officials are performing relatively well, and that shortcomings may be due to constraints at higher levels, such as the district or national government. Thus, we expect respondents assigned to the scorecard condition to⁴:

- Be more informed about the relative performance of their project in the district (manipulation check).
- Shift responsibility to lower level officials in case of poor performance.
- In the case of poor (high) performance, evaluate the performance of lower level officials more negatively (positively).
- In the case of poor (high) performance, lower (increase) their trust in lower level officials.
- Be more (less) likely to vote against the incumbents at the subcounty level in the case of low (high) performance.

⁴As secondary hypotheses, we may see increased trust in, performance evaluations of, and vote shares for district officials in the case of low performance (since respondents attribute more responsibility to lower level officials) and vice versa in the case of high performance. Furthermore we may see the increased action taken to improve spillover to other domains (*taking general action*).

- Be more (less) open to the challenger in the case of low (high) performance.
- Be more likely to take action in the case of low performance.

Second, citizens are asked to gather during the scorecard dissemination and are walked through the performance indicator of their project, thereby reducing the cost of coordination and creating common knowledge about the performance of the project. As a secondary hypothesis, we expect that this gathering lowers the cost of coordination and collective action, thus making it more likely that communities take action.

We expect the largest effects in the joint treatment condition (combining training and information).

6 Key Variables and Measurement

6.1 Dependent Variables

We group our dependent variables into two categories, those relating to the evaluation of government performance and voting behavior (A) and those relating to sense of efficacy and collective action (B). In line with the hypotheses described above, we consider dependent variables in group A the primary outcomes for the scorecard intervention, and group B to be secondary outcomes. For the training intervention, the reverse is the case: Here, we consider outcome variables in group B our primary outcomes, and those in group A as secondary outcomes.

All indicators except voting behavior will be collapsed into an averaged z-score index.⁵ As a robustness check, we will also construct indices using principal component analysis. Components are listed below each category. We consider the indices and variables under voting behavior as our main outcome variables, but will report results for all components.

6.2 Group A: Evaluation of government performance and voting behavior

1. Evaluation of government officials

(a) Village

- To what extent do you trust that your sub-project leaders make decisions in the best interest of the community? By leaders, we mean CPC or CPMC leaders.

⁵As in Kling, Liebman and Katz (2007a), we construct z-scores of each index component by subtracting the mean of the control group and dividing by the standard deviation of the control group. We then construct the index by taking the average of the z-scored components. Where necessary, index components are reoriented such that a higher value corresponds to a better outcome.

- Please tell me very honestly, if it were entirely up to you, how likely would you be to choose the same sub-project management committee (CPC or CPMC) again?
- How do you rate the performance of the LC1 Chairperson in general?
- How would you rate the performance of the CPC overall?
- How would you rate the performance of the CPMC overall?
- How would you rate the performance of the CMG members?

(b) Subcounty

- To what extent do you think your LC3 Chairperson acts in the interest of community members?
- To what extent do you think your sub-county bureaucrats such as Sub-county Chief or Community Development Officer (CDO) act in the interest of community members?
- How do you rate the performance of the LC3 Chairperson in general?

(c) District level and above

- To what extent do you think your LC5 Chairperson acts in the interest of community members?
- To what extent do you think your district bureaucrats such as District veterinarian, District engineer or program Desk Officer (NDO) act in the interest of community members?
- How do you rate the performance of the LC5 Chairperson in general?
- To what extent do you think your central government acts in the interest of citizens?

2. Voting behavior

- Voted for incumbent LC5 Chairperson or incumbent party (coded as 1 if voted for him/her, 0 otherwise)
- Voted for incumbent LC3 Chairperson or incumbent party (coded as 1 if voted for him/her, 0 otherwise)
- Voted for the ruling party
- Openness to opposition parties

6.3 Group B: Sense of efficacy and collective action

1. Sense of efficacy

- As citizens, we should be more active in questioning the actions of leaders [Agree/disagree]
- How much influence do you think you can have to make this village a better place to live?
- Some people think that as a group people in the community have no power to improve the quality of their sub-project implementation. Other people think that the community has power to improve the quality of their sub-project implementation. How much power do you think your community has?

2. Perceived collective action potential

- Agreement with the statement: Let's say there is an issue that affects many people in the community, such as a broken well, a bushy road or poor leadership. In such an instance, it is EASY to get many members of the community to come together to solve the issue, even if this takes time, effort and courage from all individuals.
- Disagreement with the statement: Let's say there is an issue that affects many people in the community, such as a broken well, a bushy road or poor leadership. In such an instance, it can be HARD to get many members of the community to come together to solve the issue, because everyone waits for someone else to do it.

3. Taking action

(a) Program-specific

- In the past year, have you YOURSELF talked to someone in your community to raise an issue regarding your sub-project, either in a meeting or one-on-one?
- Number of officials contacted for program-related complaints by respondent or others
- Whether respondent has contacted anyone to place a project-related complaint
- Whether respondent or others in the community have contacted the district or the Inspectorate of Government to place a project-related complaint
- Since the elections of 2016, how many times have you talked to members of the sub-project about the quality of project implementation

(b) General⁶

- Membership in community committee: Are you currently a member of any committee that makes decisions that affect a large portion of the community, such as a

⁶Only relevant for the testing of a secondary hypothesis of spillover effects to other domains.

committee farmers forum, a water source committee, a parish development committee, a school management committee or a boda boda association?

- In the past year, have you YOURSELF talked to someone in your community to raise an issue regarding problems with OTHER programs in your community, NOT your sub-project, either in a meeting or one-on-one?
- Number of officials contacted for non-program-related complaints by respondent or others
- Whether respondent has contacted anyone to place a non-project-related complaint
- Whether respondent or others in the community have contacted the district or the Inspectorate of Government to place a non-project related complaint
- Turnout in elections for the LC3 and LC5 chairperson

Additional analyses

As a test of our posited mechanism, we will investigate whether the treatments affect the share of respondents answering that community leaders, subproject members, or subproject leaders are primarily responsible for the quality of sub-project implementation. (Who do you see as primarily responsible for the quality of sub-project implementation?) We will also test for consistency between who is considered responsible for the performance of the project and the office targeted by complaints and changes in voting behavior. For example, we expect that people who respond that they blame their village chairperson for the outcome of the project will punish the village chairperson, not, say, the district chairperson. As a descriptive analysis, we will test whether recipient satisfaction with the project is more closely correlated with the objectively measured performance score in the information treatment.

6.4 Independent Variables

6.4.1 Heterogeneous treatment effects

- Corrupt areas, as indicated by officials
- Above/below median project performance in the district
- Partisan alignment of the subcounty and district leadership with the central government

6.4.2 Treatment Variables

- *T*: A measure that takes a value of 1 if a respondent lives in a community assigned to receive the training treatment, 0 otherwise.

- *SC*: A measure that takes a value of 1 if a respondent lives in a community assigned to receive the scorecard treatment, 0 otherwise.

6.4.3 Covariates

We will include a vector of demeaned covariates as well as their interaction with the treatment indicator (Lin et al., 2013; Imbens and Rubin, 2015) in our main specification.

- *Educ*: Years of formal education of the respondent.
- *Female*: An indicator variable that takes value 1 if the respondent is female, 0 otherwise.
- *Age*: Age in years
- *Distance*: Distance to subcounty headquarters measured in kilometers based on digital maps.
- *Plots*: Number of plots of land owned⁷

6.5 Attrition, Outliers, and Missing Values

6.5.1 Attrition

As we do not have baseline data on community members, attrition between panel rounds is not a relevant concern. However, there were cases of missing projects. When the survey team entered a village and was unable to find a project during data collection, a research assistant was sent to confirm whether the project existed. In total, 23 of the projects, or 2.6% of the sample, could not be found by the survey team at any of the data collections. For the purposes of this analysis, we will consider these projects to be non-existent. Significant efforts were made by the survey team to locate the projects and confirm their existence. It is possible that communities did not declare some projects to the survey team.

6.5.2 Outliers

We deal with outliers by capping unbounded variables at the 99th percentile of the observed values in our data.

⁷Collected post-treatment but, given the cost of land and the nature of the intervention, considered extremely unlikely to move in response to the treatment.

6.5.3 Missing covariate values

To deal with missing values on our covariates, we follow Lin, Green and Coppock (2015):

1. If no more than 10% of the covariate's values are missing, we will recode the missing values to the overall mean.
2. If more than 10% of the covariate's values are missing, we will include a missingness dummy as an additional covariate and recode the missing values to 0.

6.5.4 Missing dependent variables

To deal with missing values on our outcome measures, we will adopt the approach described in Kling, Liebman and Katz (2007b) and impute missing values by setting them equal to the mean of the respective outcome variable for the relevant treatment group.

7 Analysis

This section describes the empirical strategy we will be used to test the hypotheses generated above.

7.1 Main Analysis

We will use data from the endline individual survey to test the differential effects of each program on the outcomes presented above. We will run the following intention-to-treat (ITT) regression model:

$$Y_{ij} = \beta_0 + \beta_1 T_{ij} + \beta_2 S_{ij} + \beta_3 T_{ij} * S_{ij} + \phi R + \epsilon_{ij} \quad (1)$$

where i refers to a project or household in community j and Y_{ij} is the outcome of interest. T_i indicates whether a community was randomly selected to the social accountability training treatment, S_i indicates whether it was selected to receive the scorecard treatment, and $T_{ij} * S_{ij}$ indicates households assigned to both social accountability trainings and scorecard distribution. R is a matrix of district dummies and ϵ_i is the error term. Robust standard errors are clustered by community, the unit of randomization. The coefficient β_1 thus presents the impact of the social accountability training treatment only, β_2 the impact of the scorecard treatment only, and $\beta_1 + \beta_2 + \beta_3$ the impact of combining social accountability training and scorecard treatments.

To test for heterogeneous treatment effects, we run the following model:

$$Y_{ij} = \beta_0 + \beta_1 T_{ij} + \beta_2 T_{ij} * Het_{ij} + \beta_3 Het_{ij} + \beta_4 S_{ij} + \phi R + \epsilon_{ij} \quad (2)$$

where Het_{ij} refers to a binary variable indicating the subgroup an observation belongs to.

*In addition to the outcomes in Equation 1, we will also explore the heterogeneities discussed above using interactions with the treatment variables or by dividing the sample as necessary.

7.2 Correction for Multiple Hypotheses Testing

To correct for multiple hypotheses testing, we will use the Benjamini-Hochberg (1995) False Discovery Rate correction, defining families of hypotheses as outlined in Section 6. We will show results with the uncorrected and the corrected p-values.

8 Power Analysis

The sample size for the household survey was determined to provide the highest statistical power given a fixed budget. The intra-cluster correlation (ICC) for the main economic outcome of interest, number of cattle, is 0.045. With 574 clusters and approximately 8 people per community surveyed, the minimum detectable effect (MDE) size is below 10%. We expect similar MDEs for the political outcomes as well.

9 Timeline

Late 2014-mid 2016: Distribution of program funding and training of social action committees (median funding date late 2015)

December 2015-January 2016: Audit of community projects

January-February 2016: Dissemination of scorecards to communities

March 2016: Local government elections

June-July 2016: Household endline survey

10 Partnership and Human Subjects

International human subjects approval was obtained from Innovations for Poverty Action. Local IRB was obtained from the Uganda National Committee of Science and Technology (UNCST).

References

- Imbens, Guido W and Donald B Rubin. 2015. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.
- Kling, Jeffrey R, Jeffrey B Liebman and Lawrence F Katz. 2007a. “Experimental analysis of neighborhood effects.” *Econometrica* 75(1):83–119.
- Kling, Jeffrey R., Jeffrey B. Liebman and Lawrence F. Katz. 2007b. “Experimental Analysis of Neighborhood Effects.” *Econometrica* 75(1):83–119.
- Lin, Winston, Donald Green and Alexander Coppock. 2015. “Standard Operating Procedures: A Safety Net for Pre-Analysis Plans.” Forthcoming in *PS: Political Science & Politics*, 2016.
- Lin, Winston et al. 2013. “Agnostic notes on regression adjustments to experimental data: Reexamining Freedman’s critique.” *The Annals of Applied Statistics* 7(1):295–318.

Appendix

A Overview Page of a Sample Scorecard

Community Scorecard NUSAF2 Livestock Sub-Project

Sub-Project Name:

We have compared your sub-project with 33 other sub-projects in your district.

Your Rank

1. Health When Animals Arrived:	21 out of 33
2. Health and Animal Productivity:	22 out of 33
3. Assistance from your DVO:	7 out of 33
4. Value for Money:	20 out of 33

Overall Score: 69 out of 100

Overall Position in the District: **12 of 33**

