Agricultural Community Training and Local Governance
Experimental Evidence from Liberia

Gonne Beekman

Wageningen University, Development Economics Group

Abstract

Stimulating rural development is prominently back on the international development agenda. In this study we analyse the impact from a community training and input provision project in rural Liberia on households' livelihood status and social cohesion. We present findings from a randomized controlled trial and complement them with results from a matching procedure among a larger, quasi-experimental sample. We vary group institutions among the treatment group in a 2x2 design. We relate the project outcome to the incidence of local corruption. We find weak evidence that the project contributed to higher rice harvests and lower dietary diversity. The project did not contribute to social cohesion. Our results suggest that the project causes unintended shifts of activities within households: the project leads to an increase of time spent on the farm by children in targeted households. We find that most of our results are driven by the subgroup that received a direct democracy and leadership accountability treatment. Finally, we find suggestive evidence that local corruption undermines project impact. Our results imply that, in order to measure the full impact of a project intervention, eventual reallocations of resources must be taken into account. The interplay of project interventions with local governance structures could be a subject of further research.

Keywords: Randomized controlled trial, rural development, Liberia, direct democracy, local governance.

1 Introduction

Stimulating rural development in conflict-affected areas is a key objective for many development organisations. Civil war is sometimes called 'development in reverse' (Collier et al. 2003, p.13). War is generally believed to be destructive for both infrastructure and markets and disrupts human capital accumulation, particularly on the short term (see Collier 1999; Blattman & Miguel 2010). In addition to (temporarily) hampering economic growth, civil war is believed to erode social capital (e.g., see Colletta & Cullen 2000), and the consequences of war can last for decades.[1] Many development organisations have therefore shifted their focus from short term emergency aid to more sustainable 'reconstruction' programs. These programs are often following an 'integrated' development approach in rural areas, focussing on an array of activities that target different needs at the same time (e.g., King 2013).

We evaluate a community training project in rural post-war Liberia, implemented by an internationally operating non-governmental organisation (NGO). Many rural development programs aim to stimulate rural development either through agricultural extension services or through 'community-driven' development programs by introducing new institutions or supporting existing ones. This project aims to do both. It combines elements from a 'farmer field school' (FFS) and a 'community-driven reconstruction' (CDR) program in order to improve food-security and livelihoods, as well as strengthening social cohesion. We briefly describe both approaches below.

Farmer field schools (FFS) are a popular avenue to rural livelihood development. A number of recent papers on the livelihood impact of FFS suggest that the approach may contribute to poverty alleviation and productivity growth through improved farming techniques (e.g., Davis et al. 2012). Where the FFS approach initially aimed at increasing the adoption of specific agricultural technologies such as better pest management or the reduction fertilizer use, advocates of the FFS approach motivate that the approach should also contribute to 'educational, social and political capabilities' (Van den Berg et al. 2007). FFS could therefore rather be seen as a model for general adult learning, focussing on farmers' empowerment than as an extension service with focus on technological outcomes alone (Van den Berg et al. 2007; Friis-Hansen and Duveskog

[1] Recent empirical evidence shows that war can strengthen parochial altruism—trust and cooperation among in-group members (Gneezy & Fessler 2012; Voors et al. 2012; Bauer et al. 2014; Gilligan et al. 2010). Social cohesion between groups, however, is more likely to be weakened by civil conflict.

2012). A key characteristic of FFS is the self-selection of participants. The likelihood that FFS will be successful is largest if they are based on existing farmers groups, supported by trusted lead-farmers. For this reason, experimental evaluations of FFS are scarce. Instead, some studies employ matching techniques combined with a difference-in-differences (DD) set-up. For example, Todo and Takahashi (2013) find a large effect on agricultural income from a FFS in Ethiopia, and Larsen and Lilleør (2014) find positive effects on food security among participating households, but not on their poverty status, which may be caused by intra-household shifts of labour or consumption smoothing over time. The only randomized controlled trial on FFS we are aware of is conducted among rice farmers in China by Guo et al. (2015). The authors find mixed evidence for increase of knowledge acquisition, and effects are smaller for female and older participants.

Community-driven reconstruction (CDR) projects have increasingly gained popularity in recent years. The CDR approach is developed specifically for post-war contexts, and aims to build or reshape institutions and contribute to social reconciliation. The approach is rooted in the more general 'community-driven development' (CDD) approach, which was introduced in response to some of the shortcomings of traditional development aid programs. Mansuri and Rao (2013) estimated that the World Bank—the largest supporter of CDD programs—invested 85 billion USD on participatory development programs since the early 1990s. In the past decade the World Bank approved more than 600 CDD programs in 110 countries. These programs should be better-tailored to local needs than traditional development programs, allow for more ownership and autonomy of project participants, and welcome institutional transformation. New institutions might be adopted when local communities are exposed to democratic procedures and accountability practises. This could especially be relevant in contexts characterized by weak institutional quality. Main elements that characterize CDR projects are democratic elections of village development councils and block grants in order to carry out commity projects (see King & Samii 2014).

Notable experimental evaluations of CDR projects were carried out in the Democratic Republic of Congo (Humphreys et al. 2015), Liberia (Fearon et al. 2009), and Sierra Leone (Casey et al. 2012, discussed in greater detail below). Despite great expectations, none of these studies find robust evidence that institution building contributes to social cohesion or improved development outcomes. In fact, it is questionable whether 'new' institutions can really improve on existing ones within the brief timeline of the project activities. Newly introduced institutions are found to be ineffective if communities rely on traditional (informal) institutions (Fearon et al. 2013), and impacts of CDD programs may even be harmful if they undermine existing local institutions (King and Samii 2014).

We are aware of two studies that are related to the project we evaluate—both in terms of intervention and geographically. Casey et al. (2012) study the impact of a CDR program in Sierra Leone using a randomized experimental design. The program aims at stimulating local democracy and institutional quality through financial assistance for different types of community projects, combined with the organisation of structures that facilitate collective action (such as village development committees). Common village projects included the construction of local public goods (education, water and sanitation, etcetera), communal farming, livestock and fishing, and small business development. After completion of the project activities, the authors implemented 'structured community activities' to experimentally measure collective action and elite capture. The authors find no evidence for elite capture, perhaps because the gifts that were distributed in this context had a highly 'public' nature, and could therefore not easily be diverted. Although the authors find some positive (short term) effects on local public goods and economic outcomes, they find no evidence for (longer term) impact on collective action. A second study that is related to ours is the evaluation of the Kokoyah Millennium Villages Project in Liberia (King 2013). This project includes a large number of different interventions, the most notable being agricultural training and inputs provision and the interventions in the realm of health. The author evaluates the impact of the program on social cohesion—other outcome variables are not taken into account. She applies a DD research design, matching treatment and control groups on the village level. The results indicate that the projects slightly improved social cohesion, but that from the onset, social cohesion was not as weak as initially feared.

The project evaluation will be of particular interest to development economists and practitioners. We use a randomized controlled trial (RCT), which generates evidence on causal impacts from the project in a sample of 52 villages. As our sample is small, we complemented our analysis with quasi-experimental evidence including additional control communities to form a larger sample of 72 villages. Our analysis deploys a range of methodologies. We use a public goods game to measure cooperation and we measure leadership quality using a 'natural' field experiment which allows us to directly observe capture of project inputs by the village chief. Finally, our study rests for a large deal on detailed household questionnaires measuring livelihood indicators and time use patterns for individual household members.

Our study hopefully speaks to three literatures. First, our research speaks to the literature on rural development interventions, aiming to improve food security. We analyse whether agricultural community training contributed to rural development, and measure the project impact on livelihood indicators as well as on social cohesion. We then expand our analysis beyond assessing the mere project impact on intended outcome variables, and analyse how the intervention influences intra-household labour allocations. It is a general misconception among development practitioners that African villagers are not time-constrained. Although many villagers may not be formally employed, labour often forms a constraining factor for rural households (e.g., see Ellis 1993 for an analysis of family labour in peasant farms). Most rural households are close to full-time involved in labour-intensive farming activities. We suspect that the introduction of a development intervention that heavily relies on labour input will increase the burden on household labour. This will either happen at the expense of other activities, or lead to shifting activities to the less time-constrained individuals in the household: children. To test this hypothesis, we exploit a detailed time-use survey in order to map labour allocations among individual household members.

Second, our study speaks to the literature on community-based development. We ask to what extent newly introduced institutions on the group level mediates project performance. To this end, we introduce two institutional sub-treatments—a direct-democracy treatment and a leadership accountability treatment—in a two-by-two design. The direct democracy treatment involves project participants in the choice of training modules in a very early stage of the project. It is expected that participating in direct democracy induces democratic norms within communities (Casey et al. 2012) and as villagers are truly involved in the selection procedure, the final outcome from the election could gain legitimacy (Olken 2010). For example, Beath et al. (2012) find that in villages where projects are selected through consultation meetings, final project selection is more likely to accord to the preferences of elites than projects selected by means of a secret ballot referendum. Olken (2010) finds that under a general voting procedure, selected projects are more in favour of poor women, who might not have been involved as much in a general meeting process. More importantly, though, direct democracy leads to much more overall satisfaction among villagers with the selected project. Yet, some authors contend that involving inexperienced villagers in decisions making processes leads to better development outcomes. The argument is that established leaders can provide more technical expertise, drive, and continuity, which could lead to more productive outcomes (Bernard et al. 2010). Giving villagers means to keep their leaders accountable and the transparency of decision making processes increases. This could, in turn, lead to better collective outcomes. Previous studies have found that a higher level of monitoring indeed increase the leader’s effort as well as public goods provision (e.g., Grossman and Hanlon 2014; Olken 2007; Björkman and Svensson 2009).

Finally, our study offers a modest contribution to the literature on elite capture of project benefits and the impact from local governance quality on project outcomes in general. Few studies have empirically investigated the effects from leadership quality on project outcomes. The underlying reason is that leadership quality and project outcomes (or: economic growth) are interrelated. One notable exception on the macro-level is the cross-country analysis by Jones and Olken (2005), who demonstrate that (powerful) national leaders have a large impact on GDP in their country.[2] On the micro-level, Khwaja (2009) relates the upkeep of community projects to the project leader's quality in 99 rural communities in northern Pakistan. He finds that leadership presence positively affects a group's collective success and that this effect increases with the quality of the leader.[3] We use a direct measure of corruption, by tracking the amount project inputs captured by the village chief, which may signal local governance quality in daily life. Our measure of capture entirely coincides with the project intervention. Hence, we cannot measure the mediating effects from capture on household level project impact. Instead, we relate the incidence of capture to direct project performance indicators, which could be a prediction for the impact of the project.

We find suggestive evidence that the project contributed to higher rice harvests on farmers' private farms. The project did not contribute to social cohesion. Instead, our results suggest that social cohesion slightly decreased. Both results are driven by the subgroup that received the combination of direct-democracy (DEM) and leadership accountability (LA) treatments. The project has a robust, positive effect on time spent on farming activities by children in targeted households. We hypothesise that these children compensate for the time spent on additional farming activities by adults in the household. This effect is smallest for the groups assigned to the DEM/LA treatment combination. Finally, we find that the incidence of capture of project inputs is related with lower harvests of the project groups, controlling for the actual amount of inputs captured. This indicates that capture of public goods may negatively affect project outcomes beyond the direct negative effects from inputs diversion.

---

[2] They evaluate a sample of countries where a new national leader was installed after the sudden death (due to natural consequences or an accident) of the incumbent leader. The timing of leader replacement is thus unrelated to political factors and economic performance.

[3] Leadership quality is measured as the average of the evaluations of five community individuals (good or bad) of the project leader's quality.

The remainder of this paper is organised as follows. Section 2 presents the intervention, the research design and summary statistics of key variables. Section 3 describes the empirical strategy, and Section 4 reports results. In Section 5 we present the effect from the intervention on the allocation of time within households, and Section 6 discusses a number of mechanisms that could have affected the outcome of the intervention. Finally, Section 7 concludes.

2 Training farmers' groups in rural Liberia

2.1 The project

Our experiment evaluates an agricultural community training program targeting households in rural Liberia. The program has been implemented in two rural provinces not far from the capital city Monrovia. Infrastructure in these areas is in extremely poor condition or entirely absent. None of the rural communities, for example, is connected to the electric grid. Major livelihood activities in this region are subsistence farming and contract labour on rubber plantations, which are ubiquitous in this region. The implementing organisation is one of the many international development organisations aiming at reconstruction of Liberian society after the 14-year civil war. Most of their programs aim at stimulating food production and improving education, health care and water and sanitation. The main objective of the training project currently evaluated is to improve rural livelihoods and to stimulate food self-sufficiency through a combination of communal training and farming activities.[4] In addition, through the communal set-up the program is expected to strengthen social cohesion within communities.

These elements are deemed important in the Liberian post-war context. After the civil war that lasted for fourteen years, all major infrastructures—roads and bridges, telecommunications, power, transportation, water and sanitation systems, schools, and health facilities—had been destroyed or were neglected for years (IMF 2008). In combination with massive displacement of the Liberian population, this negatively affected income activities and undermined food security. The war is believed to have ruptured social cohesion and undermined trust (Richards et al. 2005; Ellis 2006). In this context, corruption thrives, which provides additional challenges for interventions to succeed (e.g., see Reno 1995).

[4] In a next stage of the project, which is beyond the scope of this study, the most successful training groups will be selected to continue as actual FFS.

The community training project is based on a participatory approach. Groups of about twenty farmers select their own leader from their community, who is then trained by a local development organisation. This group leader is chairing all activities that are part of the program (see the Appendix for an overview). The project contains both theoretical and experiential learning elements. In the course of four months, the group gathers in weekly meetings to discuss training modules. In addition, the group brings the newly acquired knowledge into practise on a plot of communal land, which is designated to the group by the village chief for the duration of the project. Each group receives a selection of seeds and tools for the experiential part of the training.[5] A team of six local project facilitators keeps track of the activities and each of the facilitators is expected to weekly visit eight to nine training groups, which is an intensive task. They are, however, not actively involved in the training process. The training modules and field activities last for four months, but the group is supposed to continue to tend their communal farm until crops can be harvested. After the harvest, participants can decide to continue the farming group by themselves – as long as the land is available for the group.

To test whether direct involvement of project participants matters for the project outcomes, we add two institutional sub-treatments to the existing project design, implemented in a two-by-two design (see Table 1). In the first sub-treatment, we vary the selection procedure of an additional training module, following Olken (2010). Groups were either assigned to a secret-ballot referendum procedure, wherein participants select their preferred training module through an anonymous majority vote system—referred to as 'direct democracy' (DEM), or to a traditional consensus meeting (CON). Under the consensus treatment the training module is chosen in a group meeting. Even though in the latter treatment each group member could potentially speak out, it is possible that the final outcome of the referendum is captured by a few powerful group members. Other benefits of direct democracy could be that participants can override decisions that would be in the best interest of the village elite, and the final choice outcome will generally be closer to the preference of the median voter (Matsusaka 2005).

The second sub-treatment varies whether the elected group leader can be held accountable for his performance. Half of the groups (including the group leaders) are informed that the group can replace their group leader in case they are unhappy about the leader's performance, whereas

[5] Inputs include: seeds – 25 kg rice, 5 kg corn, 3 kg beans and peanuts, 20 g pepper seed, 5 g bitter ball seed; tools – 4 cutlasses, 2 files, 4 hoes, 2 shovels, 2 watering cans.

the other half of the groups does not receive this possibility. We refer to this treatment as 'leadership accountability' (LA). We expect that, whether or not groups actually use this opportunity, the possibility alone may increase the leaders' effort. Also, as groups know they have the possibility of holding their group leader accountable, they may better monitor his performance throughout the duration of the project.

Table 1: Two-by-two design of sub-treatments

	Leader Accountability (LA)		*Total*
	Yes	No	
Direct democracy (DEM)	11	11	*22*
Consensus (CON)	11	11	*22*
Total	*22*	*22*	*44*

2.2 Research set-up

Our field experiment evaluates the impact of the training program on agricultural production, expenditures, and food. To establish a proper counterfactual, we applied a two-stage randomization design. In the first stage, sixty communities in Montserrado and Margibi—the counties were the development organisation is active—were selected. Selection of communities was bound to four conditions: (1) Communities should not have been targeted for the program before; (2) the community should be home to at least thirty households; (3) distance between selected communities should be at least five kilometres (one hour by foot, to limit spill-over effects between villages); (4) the community should be located in an area with farming potential. Communities were selected as follows: first, sixty grids of five square kilometres were randomly selected on detailed county-level maps and in each grid the most central village was chosen. These sixty villages were visited by the team of local experts, who assessed whether the village passed all four criteria. If not, the village was replaced by the next suitable village along the same road. In the second stage, sixteen households in each community were randomly selected by means of a public lottery.[6] To this end, a team of enumerators together with the village chief

[6] Letting farmers self-select into the project would probably have led to lower non-compliance rates and a more 'efficient' selection procedure. This is how the implementing development organisation normally works. However, this would not allow us to get unbiased estimates in an RCT framework. For this reason, and because the majority of the households are involved in agriculture, we deemed this choice defendable. In addition, selected farmers were free to decide whether they participated or not. In our analysis, the random treatment assignment is used as instrument for actual participation in the project.

numbered each house in the village and the numbers were transferred on lottery slips.[7] Sixteen households were randomly drawn in a public lottery. Either the household head or the spouse from selected households was eligible to participate in the research activities.

Data were collected in multiple stages. The first round of baseline data collection was conducted in April and May 2010 in fifty-two communities among 832 individuals.[8] In November and December 2010, in each of the fifty-two villages about ten additional household representatives were randomly selected according to the procedure described above, and behavioural experiments were conducted among all twenty-six individuals. Among the newly selected individuals, a short version of the household survey was conducted. Hereafter, the intervention was randomly allocated to forty-four communities.[9] We randomized treatment and control with two blocks (road and no road).[10] The project was rolled out in February 2011. In each village, twenty project participants were randomly selected from our baseline household sample. Endline surveys and experiments were conducted between January and April.

The implementation of the experiment was subject to some challenges. First, we dropped eight villages from our sample as they had received the treatment in earlier stages of the project. This left us with a very small control group of only eight communities, which increases the risk of type II errors.[11] In order to increase the sample size, we randomly selected sixteen household representatives in twenty additional control villages in April 2011, following the same selection

[7] Buildings home to more than one household would received seperate numbers for each individual household—defined as 'a group of people living under the same roof, and eating from the same pot'. When selecting the first sixteen households in each village, we selected no more than one household living in the same building. However, as this constraint restricted the number of eligible households, it was relaxed when selecting the additional ten households in each village

[8] After the first round of baseline data collection it turned out that despite our careful selection procedure eight communities had already been targeted for a program before. These communities were dropped from the sample. So instead of 60 communities and 960 individuals we remained with 52 communities and 832 individuals.

[9] The remaining eight communities serve as control communities. The research team was careful not to raise expectations about the intervention that would be rolled out in a selection of the villages at any time. The link between the research team and the implementing organization was never mentioned during the baseline research activities.

[10] Road quality is an indicator for many other village-level characteristics, such as transportation costs and food prices (Casaburi et al. 2013) and rural service delivery (Porter 2002).

[11] In an early phase of the research design we opted for a treatment group of 44 villages, and a much smaller control group of 16 villages in order to allow for sub-treatments. In each treatment bin there are 11 villages, and after dropping 8 control villages, there are 8 villages in the control group.

procedure used for the other communities in the sample. These additional twenty villages were not part of the random assignment procedure, and differ significantly from the control villages in the random sample with respect to a number of key variables (see Table 3). Hence, we present RCT results based on the randomized sample of fifty-two communities and we use PSM to reduce selection bias for results from the full sample (see Rosenbaum and Rubin 1983). Second, partly due to the many different moments of data collection, non-compliance and drop-out are high. We argue that drop-out is partly random (due to technical problems in the process of recoding identity codes between the first and second rounds of data collection), but non-compliance is not. This implies that despite of random assignment of the treatment, the sample might be subject to selection effects. We control for this using random treatment assignment as instrument for actual treatment take-up. Another implication is that the drop of observations leads to reduction of statistical power. These attrition and non-compliance are further specified in the next section.

2.3 Attrition and non-compliance

Table 2 shows the sampling frame. Panel 1 lists the initial target sample, the actual sample after baseline data collection and the difference between the two. Differences between the targeted and actual samples for treatment and control groups are small (two percent on average). As planned, exactly twenty respondents were selected for treatment in each village and attrition in the control villages is low (five percent).

Panel 2 shows participation information. This information is based on self-reported data collected among respondents in treated communities, in a survey evaluating the training group. This survey was conducted directly after the endline household survey. The data show that in treated communities both drop-out and non-compliance rates are high. First, treatment information on 28 percent of the treated sample is missing (247 observations in total). In one village, treatment information has not been collected, which explains nine percent of the attrition rate. The remaining attrition is caused by missing data from individual households in treated villages. Non-compliance is high also. Non-compliance is defined as not complying with initial allocation to treatment or control groups. From the households allocated to the treatment, 387 indeed participated in the training group (compliers), and 246 did not (non-compliers), corresponding with a high non-compliance rate of 39 percent. High attrition and non-compliance rates have several implications.

Table 2: Sampling frame

		Treated	Control	Additional control	Total
	# Villages	44	8	20	72
	# HH per village	20	26	16	
1)	Target sample	880	208	320	1,408
	Actual sample	880	201	303	1,384
	Attrition (count)	*0*	*7*	*17*	*80*
	Attrition (percent)	*0*	*3*	*5*	*5*
2)	Participated	387*	0	0	387
	Did not participate	246**	201*	303*	750
3)	Total sample at endline	633	201	303	1,137
	Attrition (count)	*247*	*0*	*0*	*334*
	Attrition (percent)	*28*	*0*	*0*	*21*

Notes: * Compliers: treated according to treatment assignment (60%). ** Non-compliers: not treated according to treatment assignment (40%).

2.4 Balance and data

2.4.1 Balance

We test whether random treatment assignment was successful using a vector of baseline-level community and household variables. Table 3 reports averages for treatment and control groups, both for the restricted RCT sample as well as for the full sample. We report averages for the treatment group, as well as eventual differences between treatment and restricted and full control groups. The last two columns report test statistics for a t-test, testing for differences between treatment and control groups.

Panel A presents our community level variables. The average community is very small, and consists of 43 households. The majority of households are involved in agriculture (68 percent). There are 22 mobile phones present in the village; hence on average, every other household owns a mobile phone (note however, that none of the villages is connected to the grid). Presence of development organizations is high: 70 percent of the villages have been targeted for development projects in the past. The size of plantations in the villages varies widely across communities. The average plantation is 154 acres whereas the largest one is 1500 acres. Finally, by design, half of the communities are located along the main road.

Panel B reports our set of household controls. Households count 4.7 members, and 13 percent of the households are female headed. Household heads are 43 years of age; compared to 39 years in the control group (the difference is statistically significant at 1 percent). 6 percent of the household heads are single and they had 2.4 years of education. Nearly all respondents are protestant (90 percent). 74 percent belong to the Kpelle tribe—the most prevalent tribe in our study region. Respondents indicate that their household owns six different assets and experienced one shock in the previous year. Finally, the war clearly had a large impact on many of our respondents. A large majority of 73 percent of the households in our sample have been displaced during the war and 30 of the households have experienced an attack.

We conclude that the randomization has been successful with respect of the restricted RCT sample, with only one out of seventeen control variables being statistically different between treatment and control groups. We must note, however, that the lack of significant differences between treatment and control groups may be caused by low power, due to the small number of observations. The full sample, including twenty additional control villages, is not balanced with respect to two community level variables, and a large number of household level variables. Villages selected into treatment have a larger share of households involved in agriculture and more NGO activity. In addition, treated households are larger, household heads are less often single, had fewer years of education and more often belong to the Kpelle tribe. Furthermore, treated households own fewer assets and experienced fewer shocks. In order to minimize the selection bias in the full sample, we turn to a PSM framework (Rosenbaum and Rubin 1983). We estimate the propensity score based on this set of unbalanced control variables, using nearest neighbour matching with replacement. Next, we estimate our regression models including frequency weights based on the weights assigned to the control variables in the PSM procedure, as well as a set of household and community level covariates.

2.4.2 Description of outcome variables

We measure the impact of the project on harvest of the two major staple crops: rice and cassava. These crops are grown by the majority of farmers. In order to measure the effect on daily farm practise, we only measure the harvest on farmers' private farms (thus not taking into account the harvest from the communal project farm). Rice seed was the most important input provided by the project. It is expected that rice harvest increases after learning more about better rice farming techniques, as farmers can directly apply the newly acquired knowledge on their private farms. We include effects on cassava harvest to test for potential indirect effects from the intervention. Next, we measure the impact from the project on various income indicators: (*i*) the household's two-weekly expenditures on food and non-food items and (*ii*) a household dietary diversity score (HDDS).[12] Panel A in Table 4 provides descriptive statistics of our set of key outcome variables for the treatment group and for both the restricted and full controls groups separately.

We also measure the effects from the program on cooperation and trust. The project aims to improve cooperation and trust by letting farmers work together. It is hoped and expected that farmers will recognise the benefits of cooperation in the training group, and that cooperation and trust among community members is supported, also in daily life. We measure social cooperation using the results from a simple public goods game (PGG) that we conducted in a random subset of the village sample. In the PGG participants were grouped in groups of four players, and asked to allocate five tokens to a public or to their private account. Each token kept in the private account was worth 10 LD to the individual player, and each token shared in the public account was worth 5 LD to each of the players (so 20 LD in total). They played the game for five rounds, in changing group compositions, to allow for learning. Households assigned to the treatment group contributed 1.5 tokens in the fifth round, which is slightly more than the contribution in the control group in the RCT sample, but not different from the full control group in the PSM sample. Our trust variable is measured in the household survey on a scale from 1 to 5, where 1 refers to very little, and 5 refers to a lot of trust in fellow community members (see Panel A in Table 4).

[12] The HDDS is defined by the number of items a household consumed from twelve different food categories in a certain reference period (two weeks, in our case). This measure is seen as a good predictor for nutritional status, especially for children (see Swindale and Bilinsky 2006; Arimond and Ruel 2004).

Table3: Balance test key variables (at baseline)

	TREATMENT GROUP					CONTROL RCT SAMPLE					CONTROL PSM SAMPLE					RCT	PSM
	Obs	Mean	SE	Min	Max	Obs	Mean	SE	Min	Max	Obs	Mean	SE	Min	Max	p-value t-test	p-value t-test
Panel A: Community variables																	
# Households	41	43.34	5.43	3	145	8	41	13.29	8	127	28	31.93	5.42	8	127	0.86	0.16
Share of agri hh's	42	0.68	0.03	0.15	1	8	0.71	0.04	0.5	0.85	28	0.56	0.04	0.5	0.85	0.59	0.04
# GSMs	41	21.78	5.47	2	200	7	8.57	1.81	3	18	20	17.5	5.19	2	95	0.33	0.62
NGO activity (b)	44	0.70	0.07	0	1	8	0.63	0.18	0	1	27	0.44	0.10	0	1	0.66	0.03
Plantation farmed (acre)	44	154.14	42.41	0	1500	8	32.75	16.92	0	110	27	54.15	36.94	0	1000	0.23	0.11
Main road (b)	44	0.5	0.08	0	1	8	0.5	0.19	0	1	28	0.57	0.10	0	1	1.00	0.56
Panel B: HH variables																	
Hh size	856	4.74	0.07	1	15	172	4.44	0.17	1	13	475	3.67	0.09	1	13	0.10	0.00
Female head (b)	739	0.13	0.01	0	1	154	0.08	0.02	0	1	452	0.12	0.02	0	1	0.11	0.71
Age	850	42.50	0.51	16	90	171	39.48	1.11	17	86	471	41.13	0.66	17	86	0.01	0.10
Single (b)	880	0.06	0.01	0	1	173	0.08	0.02	0	1	476	0.11	0.02	0	1	0.62	0.00
Years of education	877	2.38	0.14	0	16	173	2.53	0.30	0	12	475	3.80	0.22	0	19	0.65	0.00
Protestant (b)	845	0.90	0.01	0	1	170	0.93	0.02	0	1	470	0.90	0.02	0	1	0.28	0.68
Kpelle (b)	851	0.74	0.01	0	1	170	0.75	0.03	0	1	470	0.69	0.03	0	1	0.78	0.03
Assets	871	6.00	0.09	0	13	175	6.02	0.20	0	12	469	6.55	0.13	0	15	0.93	0.00
Displaced (b)	839	0.73	0.02	0	1	168	0.73	0.03	0	1	469	0.74	0.02	0	1	0.96	0.89
Shocks	880	1.03	0.04	0	4	168	1.08	0.08	0	4	476	1.24	0.08	0	4	0.57	0.00
War attack	822	0.30	0.02	0	1	169	0.26	0.03	0	1	447	0.33	0.03	0	1	0.31	0.31

Notes: The bold variables are included as matching variables in the PSM procedure. All variables that significantly differ between treatment and control groups are included as control variables in the PSM regression models.

Table 4: Descriptive statistics outcome variables (at endline)

	TREATMENT GROUP					CONTROL RCT SAMPLE					CONTROL PSM SAMPLE					RCT p-value	PSM p-value
	Obs	Mean	SE	Min	Max	Obs	Mean	SE	Min	Max	Obs	Mean	SE	Min	Max	t-test	t-test
Panel A: Key outcome variables																	
Rice harvest	181	874.54	178.51	0	3750	43	242.47	113.23	0	20000	119	658.34	162.17	0	12500	0.09	0.40
Cassava harvest	93	3.94	1.49	0	100	23	2.26	1.38	0	30	49	1.53	0.73	0	30	0.59	0.26
Food expenditures	662	2563.10	61.01	0	6710	151	2562.76	120.47	0	12550	445	2481.18	65.91	0	6710	1.00	0.37
Non-food expenditures	661	2236.70	130.61	0	15150	151	2008.80	210.46	0	25425	444	1757.35	106.66	0	16350	0.43	0.01
Dietary diversity score	658	4.99	0.07	0	10	151	5.16	0.15	0	9	445	5.45	1.94	0	10	0.30	0.00
PGG contribution	339	2.72	0.09	0	5	119	2.73	0.16	0	5	237	3.36	0.11	0	5	0.95	0.00
Trust comm. members	647	3.64	0.04	1	5	148	3.76	0.07	1	5	440	3.85	0.04	1	5	0.16	0.00
Panel B: Time use variables																	
Farming head	588	3.22	0.19	0	24	133	3.20	0.37	0	11.75	414	3.05	0.22	0	18.5	0.97	0.57
Tapping head	588	1.16	0.13	0	12	133	1.76	0.29	0	10.50	414	1.56	0.16	0	12	0.05	0.05
Housework head	588	1.53	0.11	0	24	133	1.25	0.18	0	9.5	414	0.94	0.09	0	11	0.26	0.00
Recreation head	588	0.38	0.05	0	10.5	133	0.41	0.10	0	9	414	0.27	0.05	0	10	0.78	0.17
Farming spouse	282	2.31	4.18	0	19.75	61	2.40	0.53	0	12.75	211	1.33	0.23	0	12.75	0.87	0.01
Tapping spouse	282	0.50	0.13	0	15	61	0.71	0.29	0	10.75	211	0.20	0.09	0	10.75	0.51	0.09
Housework spouse	282	2.97	0.15	0	16.25	61	3.05	0.35	0	12	211	3.16	0.17	0	13.5	0.83	0.41
Recreation spouse	282	0.50	0.09	0	13	61	0.34	0.15	0	7	211	0.28	0.08	0	8.5	0.45	0.09
Farming child	145	1.36	0.25	0	16	23	0	0	0	0	56	0.58	0.30	0	10	0.03	0.08
Schooling child	145	4.1	0.37	0	24	23	4.25	0.80	0	0	56	4.30	0.45	0	10	0.88	0.76
Recreation child	145	1.89	0.21	0	12.5	23	2.88	0.70	0	0	56	2.50	0.42	0	11.75	0.10	0.16

3 Empirical strategy

To probe the impact of training program on food security and social cohesion, we do two things. First, we calculate intent-to-treat effects (ITT) from the overall intervention, as well as from each of our four individual sub-treatments (DEM/LA, DEM, CON/LA and CON). To this end, we measure the effect of the treatment assignment z regardless of actual treatment uptake $d(z)$.[13] ITT is defined as follows:

$$ITT_{i,D} = Y_i\big(d(1)\big) - Y_i\big(d(0)\big)\,, \qquad (1)$$

where the *ITT* for individual *i* is the difference between outcome *Y* for individuals assigned to the treatment *(d(1))* and outcome *Y* for individuals assigned to the control group *(d(0))*. Because the treatment was randomly assigned to subjects, treatment assignment is exogenous, as confirmed by balance tests in Table 2. Any observed effect on *Y* can hence be attributed to the treatment. To calculate the ITT, we estimate a simple OLS regression model:

$$Y_{ij} = a + \beta_{ij} * z_{ij} + \varepsilon_{ij} \qquad (2)$$

where Y_{ij} is the ITT of the outcome variable, a is the intercept $E[Y_{ij}(d(0)]$, estimating the outcome for untreated individuals i in community j, β is the treatment effect of treatment assignment z, and standard errors ε are clustered on the community level (j=1,...,52).

The sampling frame in Table 2 shows that treatment assignment often did not coincide with actual treatment take-up. Non-compliance amounts to about forty percent, indicating that forty percent of the individuals in our sample either did not participate in the training group although they were selected for treatment. Hence, the ITT effects likely underestimate actual treatment effects (assuming that treatment effects are larger for participants than for non-participants). To estimate the effects of the program on those individuals who were actually treated, we also estimate the local average treatment effects (LATE), the effect of the treatment on compliers:

$$LATE_{i,D} = Y_i\big(d(1)=1\big) - Y_i\big(d(0)\big)\,, \qquad (3)$$

where the *LATE* for individual *i* is the difference between outcome *Y* for treated individuals *(d(1)=1)* and outcome *Y* for non-treated individuals *(d(0))*.

[13] The treatment allocation has three potential outcomes. Households assigned to the treatment are actually treated (*d(1)=1*), and households assigned to the control group are untreated (*d(0)=0*). These are the compliers. Then, households assigned to the treatment may opt-out (*d(1)=0*). These are non-compliers.

As actual participation in the program is endogenously determined (the likelihood that someone will decide to join is affected by certain personal characteristics) we run an instrumental variable (IV) model:

$$Y_{ij}=a+\beta_1 d_{ij}^{*}+\varepsilon_{ij} \tag{4a}$$

$$d_{ij}^{*}=\theta+\gamma_1 z_{ij}+\epsilon_{ij} \tag{4b}$$

where the endogenous treatment status d^* is instrumented by the exogenous treatment assignment status z.

We estimate the ITT and LATE of the treatment on harvest, livelihood status and food security, and cooperation and trust. As the results from the RCT analysis are based on a small sample, which increases the risk of type II errors, we also consider the full sample including the twenty additional control villages that were selected after the implementation of the project. In what follows, we present results for the restricted RCT sample alongside results from the full sample, using a PSM framework.

4 Empirical results

Table 3 presents the ITT (Panel A) and LATE (Panel B) on harvest of the two major staple crops: rice and cassava. The ITT on rice harvest, presented in column (1) is large and highly significant (an increase of 632 rice bundles harvested per household in the treatment group compared to an average of 242 bundles in the control group). As expected, the LATE are even larger (834 bundles). This implies that the project had a large, positive effect on rice harvest, regardless whether people actually participated in the project (an increase of 260 percent among villagers assigned to the treatment, and an increase of 320 percent among actual project participants).[14] The results are driven by a small number of relatively large observations. We suspect that some of these observations may be outliers, are these farmers are not characterized by other specific household characteristics, such as higher expenditures or consumption patterns. In columns (2) and (3) we present results from regression models cropping the top 1 and 5 percent of the observations (represented by only 2 and 11 observations). The effect is much smaller in column (2), and disappears in column (3). The results for the full sample in column (5)

[14] Note that the size of these figures should be interpreted with caution: crop harvest has been measured with some of measurement error due to use of many different units that were ex-post transposed to a single unit—i.e. bundles in the case of rice, and bags in the case of cassava.

are also positive but insignificant. Finally, the treatment has no effect on cassava harvest [see columns (4) and (6)].

Zooming in on the effects from our sub-treatment in Table 4, it seems that the ITT effect is entirely driven by groups assigned to the direct-democracy and leadership accountability treatment (DEM/LA). We test whether coefficients from sub-treatments differ from each other; corresponding p-values are reported in the bottom panel. The effect sizes do not significantly differ across the sub-treatments. The DEM/LA treatment combination only has a significantly larger effect on rice harvest than the DEM treatment alone in column (2), excluding the top 1 percent of observations. As before, the result disappears when cropping the top 5 percent of observations.

Our results suggest that participation in the training group may have a positive effect on rice harvest on people's private plots, as people might directly apply newly acquired knowledge about farming strategies on their own land, too. The effect is driven by groups where members were directly involved in the project design and who could keep their group leader accountable for his performance. The size of the effect should be interpreted with caution, however, as it is possibly partly driven by a small number of outliers. The absence of an effect on cassava harvest indicates that this result is not driven indirect mechanisms, other than participation in the training.

Table 3: Harvest

	RCT				PSM	
	Rice harvest *All obs.*	Rice harvest *Top 1% obs. excluded*	Rice harvest *Top 5% obs. excluded*	Cassava harvest	Rice harvest	Cassava harvest
	(1)	(2)	(3)	(4)	(5)	(6)
ITT	632.1***	446.31**	131.31	1.675	79.29	2.128
	(215.2)	(172.61)	103.13	(1.936)	(272.4)	(1.959)
N	224	222	213	116	347	188
R^2	0.013	0.014	0.004	0.003	0.031	0.058
LATE	870.02***	605.09***	163.38	1.304	128.1	2.679
	(288.80)	(226.53)	(140.54)	(2.470)	(360.1)	(2.568)
N	217	215	206	111	341	184
R^2					0.030	0.060

Notes: Standard errors are clustered at the village level and reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Regression models in columns (5) and (6) include household controls (household size, marital status, years of education, ethnicity, assets, shocks, displacement during war) and community controls (share of households involved in agriculture, NGO activity). Rice harvest is measured in bundles, cassava harvest in bags.

Table 4: Sub-treatment effects on harvest (ITT)

	Rice harvest	Rice harvest	Rice harvest	Cassava harvest

	All obs.	*Top 1% obs. excluded*	*Top 5% obs. excluded*	
	(1)	(2)	(3)	(4)
1. DEM/LA	1002.2**	1002.2**	226.4	4.822
	(420.1)	(420.1)	(159.9)	(4.729)
2. DEM	312.4	23.47	23.47	-0.511
	(360.0)	(129.8)	(129.9)	(1.511)
3. CON/LA	322.0	322.0	187.6	1.906
	(200.6)	(200.7)	(172.8)	(3.759)
4. CON	828.5*	442.2	125.6	1.160
	(447.9)	(273.7)	(147.5)	(2.415)
N	224	222	213	116
R^2	0.028	0.062	0.012	0.022
1=2	0.21	0.03	0.24	0.26
1=3	0.14	0.14	0.85	0.62
1=4	0.77	0.26	0.59	0.47
2=3	0.48	0.16	0.37	0.51
2=4	0.36	0.14	0.52	0.45
3=4	0.29	0.71	0.75	0.86

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The bottom panel reports p-values for tests that test whether parameters are equal.

In Table 5 we estimate the treatment effects (ITT and LATE) of the training program on expenditures and dietary diversity. We find no statistically significant effects from the program on any of our livelihood indicators in the restricted sample in columns (1)-(3). The lack of results is due to the large standard errors combined with a limited number of observations. In the full PSM sample in columns (4)-(6) we only find a marginally significant, negative LATE for the household dietary diversity score (the ITT is also negative, but not statistically significant). Our results provide no more than suggestive evidence that the project might lead to a slightly lower dietary diversity score. Perhaps, by relying on few nutritious food items, households free up resources for expenditures on schooling, clothing and other non-food items.

Table 5: Expenditures and dietary diversity

	RCT			PSM		
	Food expenditures	Non-food expenditures	HDDS	Food expenditures	Non-food expenditures	HDDS
	(1)	(2)	(3)	(4)	(5)	(6)
ITT	0.338	227.902	-0.173	74.826	166.163	-0.401
	(190.744)	(263.878)	(0.242)	(142.825)	(282.995)	(0.253)
N	813	812	809	1276	1274	1272
R^2	0.000	0.001	0.001	0.087	0.071	0.054
LATE	-26.933	417.683	-0.285	71.996	283.421	-0.672*

	(304.959)	(437.798)	(0.391)	(222.493)	(460.923)	(0.402)
N	779	778	775	1247	1245	1243
R^2				0.095	0.071	0.046

Notes: Standard errors are clustered at the village level and reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Regression models in columns (4)-(6) include household controls (household size, marital status, years of education, ethnicity, assets, shocks, displacement during war) and community controls (share of households involved in agriculture, NGO activity).

Table 6 reports the effects of the intervention on social cohesion. We find no effects from the intervention on contributions in the public goods game, or for reported trust in the restricted sample [columns (1)-(2)]. However, nearly all coefficients are negative. Considering the full PSM sample, we find a marginally significant negative ITT for contributions in the PGG in column (3): contributions in the public goods game decrease by 0.6 tokens (equivalent to 18 percent relative to the control group). Zooming in on our sub-treatments in Table 7, we find that the negative effect is driven by the DEM/LA treatment combinations: assignment to this sub-treatment reduces contributions in the public goods game by 0.8 tokens (equivalent to a reduction of 30 percent relative to the control group). The ITT is significantly different from the other tree sub-treatments. Our results indicate that the project did not manage to stimulate social cohesion in the best case and that in the worst scenario the intervention might have undermined social cohesion, especially in those groups where participants are most intensively involved in decision making and monitoring. We speculate that members in these groups may have had higher expectations from the project impact than group members in other sub-treatments, and 'learned' that cooperation with the group does not pay-off as much as expected.

Table 6: Social cohesion

	RCT		PSM	
	PGG contribution (Round 5)	Trust in community members	PGG contribution (Round 5)	Trust in community members
	(1)	(2)	(3)	(4)
ITT	-0.011	-0.075	-0.613*	-0.145
	(0.439)	(0.127)	(0.357)	(0.121)
N	458	790	789	1250
R^2	0.000	0.001	0.090	0.052
LATE	0.049	-0.062	-1.050	-0.184
	(0.812)	(0.202)	(0.666)	(0.189)
N	424	759	761	1223
R^2			0.070	0.046

Notes: Standard errors are clustered at the village level and reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Regression models in columns (4)-(6) include household controls (household size, marital status, years of education, ethnicity, assets, shocks, displacement during war) and community controls (share of households involved in agriculture, NGO activity).

Table 7: Sub-treatment effects on social cohesion (ITT)

ITT	(1) PGG contribution (Round 5)	(2) Trust in community members
1. DEM/LA	-0.836*	-0.099
	(0.470)	(0.196)
2. DEM	0.232	-0.011
	(0.504)	(0.187)
3. CON/LA	-0.068	-0.229
	(0.497)	(0.168)
4. CON	0.694	0.016
	(0.422)	(0.145)
N	458	790
R^2	0.082	0.011
1=2	0.01	
1=3	0.05	
1=4	0.00	

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The bottom panel reports p-values for tests that test whether parameters are equal.

5 Allocation of time

Farming in Liberia is, like in much of sub-Saharan Africa, a particularly physically demanding activity. It involves clearing land of shrubs and bushes and 'digging up the soil' with a hand hoe to make the land suitable for cultivation (Wairimu et al. 2014). Land is usually used for one or more farming seasons, after which it is left fallow for the next farming season. As farming activities are mostly performed using not more than small hand tools, farming—the major livelihood activity of most rural households—is extremely labour intensive. This is at odds with the increasing focus of public works based programmes, that heavily rely on precisely the scarcest resources in communities: labour (e.g., Wairimu 2014). Activities on the communal farm are additional to the private farming activities that most households undertake. Those households not involved in farming are either not able to perform physically demanding labour, or they are involved in other, non-farm income activities. This means that asking households to participate in labour intensive community-projects will increase the strain on available labour in the household. Consequently, project participants may shift labour input from their private farm to the communal project farm, reduce time spent on other activities, or involve other household members in farming activities. On being asked how they compensated for the time they spent on the project training modules and field work, nearly half of the project participants respond that they abandoned their regular activities. Only 14 percent of the respondents indicate that someone else took over their activities, and 35 percent respond that they worked more or had less time for leisure (see Table 8).

Table 8: Compensation for time spent on SGP activities

	Obs	Mean	S.D.	Min	Max
Abandoned regular activities	424	0.47	0.50	0	1
Someone took over my work	424	0.13	0.35	0	1
Worked more	424	0.22	0.42	0	1
Hired labour	424	0.02	0.18	0	1
Less leisure time	424	0.13	0.38	0	1

To test whether the intervention caused shift in time use between activities or between household members, we collected detailed information on the amount of time spent on various activities by all household members using a time use survey measuring all activities performed on a single day.[15] We assess the treatment effects of the intervention on time use for the most important activities for the household head, spouse, and children: farming, rubber tapping, house work and recreation for household head and spouse, and farming, schooling, and recreation for children. Panel A in Table 9 reports the ITT and LATE for the restricted RCT sample. We find no effect from the intervention on the amount of time spent on any of these daily activities by the household head or spouse. However, the treatment leads to children spending significantly more time on the farm than their peers in the control group. The ITT is 1.4 and the LATE 2.1, compared to an average of zero in the control group.[16] This implies that the treatment leads to an increase of 1.4 hours spent on the farm by children in households assigned to the treatment, and by 2.1 hours by children in households that were actually treated. This result is replicated in the full PSM sample in Panel B: the ITT is 0.7 and the LATE 1.1, which corresponds to a huge

[15] In the time use survey respondents were asked to meticulously report each activity performed by each household member in the course of the last regular working day (mostly 'yesterday'), including starting and ending time, beginning from 4 a.m. If present, both head and spouse were interviewed separately. If not, the interviewee would estimate time use activities for the other household members.

[16] Note that the number of observations in the control group is small. Indeed, none of the 23 households in the control group indicated that their children worked in the farm on the last working day.

increase of 120 and 190 percent, respectively. These effect sizes should be taken with a pinch of salt, given the small number of observations both in the restricted sample, as well as in the full sample. We suspect that increased time spent on the farm by children is related to less time spent on other major activities. Coefficients for time spent on schooling and recreation are indeed negative, but not significant. Perhaps the effects are not picked up as time spent on the farm is compensated by less time spent on a range of many different activities.

Zooming in on the mediating effects from our sub-treatments, we find that the increase in time spent on the farm is significantly smaller in training groups assigned to the DEM/LA sub-treatment: children from households in this sub-treatment spend only 0.4 hours more on the farm, compared to an increase of 1.5 and 2 hours in groups assigned to the direct democracy or consensus treatments (both without possibility to hold the leader accountable). Note that the ITT in the other three sub-treatments do not differ significantly among each other. We speculate that the maximal involvement of group participants may raise their awareness of the importance of children's education, which might weaken the effect of reallocation of labour to children. Under the traditional consensus meeting (CON sub-treatment), the increase of time spent on farming by children seems to be compensated by a similar reduction of time spent on schooling. Conversely, the coefficient for time spent on schooling is positive (but insignificant) for the DEM/LA sub-treatment.

Our results suggest that net time spent on farming by adults remains unchanged. Most people probably simply shift some of their farming activities from their private farm to the project farm. If people spend more time on the communal plot, communal production might increase, whereas production on the private plot would probably decrease. If this assumption holds, then we can only expect that the project is successful in terms of increasing household level productivity if per-household productivity on the communal farm is greater than productivity on the private farm. Unfortunately, we lack the data to assess land productivity of private versus project plots, so we cannot test this hypothesis. One caveat should be taken in mind. The time use data were collected a couple of months after the end of the main farming season, meaning that we measure the shifts in time use that *persist* after a couple of months. By the time the endline survey data were collected many farmers had probably reverted to their regular farming activities. Had we performed the time use survey during the farming season, we would likely have found much larger effects on labour input and labour shifts among household members.

Table 9: Time use

	Head				**Spouse**				**Children**		
	Farming	Tapping	Housework	Recreation	Farming	Tapping	Housework	Recreation	Farming	Schooling	Recreation
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Panel A: RCT Sample											
ITT	0.014	-0.593	0.275	-0.034	-0.0949	-0.210	-0.079	0.158	1.357***	-0.150	-0.994
	(0.531)	(0.552)	(0.297)	(0.142)	(0.857)	(0.445)	(0.389)	(0.250)	(0.214)	(1.269)	(0.874)
N	721	721	721	721	343	343	343	343	168	168	168
R^2	0.000	0.005	0.002	0.000	0.000	0.001	0.000	0.002	0.027	0.000	0.016
LATE	0.028	-0.931	0.510	-0.063	-0.007	-0.329	-0.170	0.189	2.116***	-0.349	-1.559
	(0.857)	(0.890)	(0.483)	(0.230)	(1.438)	(0.743)	(0.657)	(0.415)	(0.378)	(1.866)	(1.286)
N	691	691	691	691	325	325	325	325	161	161	161
Panel B: PSM Sample											
ITT	-0.097	-0.483	0.358	-0.017	0.366	-0.018	0.273	0.222	0.734*	0.257	-0.503
	(0.574)	(0.513)	(0.238)	(0.195)	(0.590)	(0.306)	(0.265)	(0.227)	(0.414)	(0.922)	(0.645)
N	1116	1116	1116	1116	589	589	589	589	218	218	218
R^2	0.019	0.031	0.044	0.021	0.086	0.080	0.049	0.024	0.085	0.083	0.076
LATE	-0.188	-0.762	0.633*	-0.039	0.838	-0.001	0.434	0.311	1.132**	0.319	-0.881
	(0.915)	(0.819)	(0.380)	(0.308)	(1.006)	(0.540)	(0.464)	(0.379)	(0.565)	(1.249)	(0.864)
N	1091	1091	1091	1091	572	572	572	572	212	212	212
R^2	0.018	0.025	0.033	0.021	0.088	0.080	0.042	0.026	0.066	0.080	0.059

Notes: Clustered standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All regression models include household controls (household size, marital status, years of education, ethnicity, assets, shocks, displacement during war) and community controls (share of households involved in agriculture, NGO activity).

Table 10: Sub-treatment effects on children's time-use (ITT)

	(1) Farming	(2) Schooling	(3) Recreation
1. DEM/LA	0.398**	1.234	-0.763
	(0.183)	(1.417)	(1.000)
2. DEM	1.494***	-0.415	-0.399
	(0.465)	(1.276)	(0.961)
3. CONS/LA	1.278**	2.241	-1.667*
	(0.497)	(1.369)	(0.908)
4. CONS	1.961***	-2.328*	-1.297
	(0.278)	(1.294)	(0.917)
N	168	168	168
R^2	0.062	0.138	0.043
1=2	0.03	0.07	0.60
1=3	0.10	0.32	0.14
1=4	0.00	0.00	0.39
2=3	0.75	0.00	0.02
2=4	0.39	0.01	0.11
3=4	0.24	0.00	0.42

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The bottom panel reports p-values for tests that test whether parameters are equal.

6 Capture of project inputs

Community leaders can play a mediating role in the failure or success of development interventions via two main channels. First, community leaders may play a direct mediating effect by channelling away project inputs for their own benefit (e.g. Platteau and Gaspart, 2003; Platteau, 2004). Second, good leadership may (indirectly) create an enabling environment for new initiatives to succeed. As community leaders fulfil a key position in many African societies, often based on traditional authority, their support is crucial for the potential success of development interventions (e.g., see Kyamusugulwa & Hilhorst 2015).

We measured the quality of leadership by the amount of 'capture' of program inputs by local leaders. Just before the start of the training program, project inputs (seeds and tools) were delivered to the 44 treated communities, and community leaders were asked to store them in their private hut for three days. After three days, a field worker would publicly weigh the seeds and officially hand over the inputs to the project beneficiaries. If any of the project inputs was missing after these three days, we define the leader as 'corrupt'. According to this (rather narrow) definition, 22 of the community leaders are corrupt, and 22 are not. As our corruption measure coincides with the treatment (we have no measure for corruption in control communities), we cannot analyse the interaction between the project implementation and the incidence of capture.

Instead, we relate capture by the village chief to the performance of the training groups. We define 'group performance' in terms of two different variables: the number of active members and whether the group harvested rice at the end of the growing season.[17] We argue that both participation in the group, as well as whether the group actually harvested, could function as predictors for eventual project impact. Questions about group performance were asked to all project participants, and then averaged at the community level. Groups were reported to have sixteen active members on average (ranging between four and twenty-five members). Also rice harvest rates vary greatly: only 50 percent of the group members indicate they harvested rice (see Table 11).[18]

Table 11: Elite capture and SGP group performance

	Obs	Mean	S.D.	Min	Max
Capture (b)	44	0.48	0.51	0	1
Active members	43	15.85	4.44	4.33	25.19
Rice harvested (b)	43	0.50	0.37	0	1

Notes: (b)=binary variable

We run the following regression model:

$$P_j=\beta_1 Q_j+\gamma X_j+\varepsilon_j \qquad ,$$

where project performance P in community j is explained by the quality of local leadership Q, a binary variable taking the value 1 if the village leader diverted some of the project inputs (j=1,...,44). X is a vector of community level controls and ε is the error term.

In Table 12 we report results of regressions of group performance on capture of project inputs. We find no evidence that leadership quality matters for the number of active members in column (1). However, our result in column (2) indicates that input diversion is negatively correlated with rice harvest. We control for the direct effects of input diversion (if project inputs are stolen there is obviously less seed to plant and the change of harvesting the crop is smaller) by including the

[17] Rice is the major staple crop in Liberia, and the most important provided to all groups. Apart from rice, groups also received a variety of vegetable seeds, but these seeds were provided in much smaller quantities and were often not harvested as seeds did not always germinate.

[18] We also have data for the actual quantity of the harvest. However, these variables are measured with a lot of measurement error due to conversion from a multitude of local units to one standard unit.

amount of rice seed diverted. The coefficient indicates a decrease of 28 percentage points. Our results show that the decreased likelihood of rice harvesting vis-a-vis a corrupt chief is not a direct consequence of input diversion (it does not matter how much rice seed was diverted; in fact, rice seeds were not stolen much at all). Instead, we hypothesize that input diversion by the chief signals his general support for 'community-goods': in communities with a supportive chief, projects may stand a better chance to succeed. This relation may be influenced by unobserved effects, such as shared village norms, that both affect group performance and leadership quality. Nevertheless, our result suggests that 'good' leadership may indeed provide an enabling environment for projects to succeed—beyond the direct effect of input provision. Our result is in line with the results in Beekman et al. (2013), who find that 'corrupt' leadership also leads to lower rice harvests on people's private farms.

Table 12: Elite capture and group performance

	Active members	Rice harvest
	(1)	(2)
Panel A: Corruption		
Inputs missing (b)	-1.598	-0.264**
	(1.473)	(0.127)
Rice missing (grams)		0.00004
		(0.00004)
Constant	11.153	0.382
	(14.212)	(1.100)
Controls	Yes	Yes
N	41	41
R^2	0.309	0.424

7 Conclusions and discussion

Community participation, especially in the setting of rural reconstruction after a period of civil war, is seen as a viable way to escape poverty by many development agencies (e.g., Burde 2004). In this paper, we evaluated a rural community-based training program in post-war Liberia that combines elements from farmer field schools and community-driven reconstruction. We apply a randomized controlled trial, complemented with results from a larger, non-random sample, based on a matching procedure.

Our results indicate that the project has contributed to higher rice harvest—the main staple crop in Liberia. However, the effect size needs to be interpreted with caution. The treatment effect estimator is imprecise as it is measured with relatively much noise. Furthermore, our study is based on a relatively small number of observations: the project was carried out in forty-four communities, and the number of control communities is much smaller. This increases the risk of type II errors due to low power. We re-estimated our models using a larger, non-random sample of villages. Although we reduced the estimation bias using a propensity score matching procedure, this may not rule out selection bias based on unobservable effects (see Angrist & Pischke 2008). The effects we do find, seem to be driven by training groups allocated to a direct-democracy and leadership-accountability treatment. We also find that this sub-treatment leads to lower willingness to cooperate in the aftermath of the project. Perhaps, members in these groups may had high expectations from the project impact, and experienced that cooperation with the group does not pay-off as much as expected. Finally, results suggest that an 'enabling institutional environment' may matter for direct project outcomes: group in villages with a stealing chief are less likely to harvest their crops.

One mechanism that may explain the limited impact is the possible reallocation of labour within the household. Like all labour-intensive project interventions for the rural poor—'cash-for-work' is another example—this project is based on the assumption that labour is an abundant production factor. This assumption is probably wrong. We observed that farming is highly labour intensive, and people are spending most of their time on the farm or on other income activities. Those people who are not working are not fit for physical labour. As the community project targets exactly the scarcest production factor—labour—it is likely that households shifted labour from their private to the public farm, while total farming time remained unchanged. This is exactly what we find, based on results from a time-use survey. In addition, we find evidence that the intervention leads to higher involvement of children in farming activities, probably partly at expense of their time spent on schooling. This does, however, not lead to higher harvests or improved livelihood status. A relevant question to further explore in the context of this project is whether communal farming is more (or less) efficient than farming on a private plot. Whereas there may be scale-advantages—sharing inputs and knowledge, these are probably overruled by disadvantages such as free-riding behaviour. Even if each farmer would contribute maximally, we may wonder whether the resulting communal output would be higher than the sum of individuals' output from each of their private farms. Unfortunately, we lack detailed data on productivity of the communal farms vis-a-vis the private plots, so this assumption remains untested. The key lesson, however, is that if households are labour constrained, then introducing new activities will inevitably lead to shifts in labour allocation within the household. Hence, a newly introduced intervention will only pay off when these shifts lead to more productive allocation of labour.

Our dataset is subject to a high attrition rate. A significant share of control villages had to be dropped as they had received the treatment before. We did select additional control villages in order to expand the control group, but, unfortunately, only *after* the random treatment assignment. The newly collected villages were thus not part of the experimental framework. In addition, the sample is subject to a high attrition rate on the household level. A major reason for attrition between baseline and endline data collection is that households received different identification codes after the first round of data-collection. In the process of recoding that followed, a number of observations were lost. The implication is that the control group in the experimental framework is very small, leading to low power and increased chance of Type-II errors, or 'false negatives'. In other words, we might fail to reject the null hypothesis when it is actually false. This means that our results might underestimate the actual effects from the program. For this reason, we expanded our control group with the additionally collected control villages, using a matching strategy. Yet, propensity score matching is no panacea either, as it remains a control strategy (Angrist & Pischke 2009). 'Since the core assumption underlying causal inference is the same for [matching and regression], it's worth asking whether or to what extent matching really differs from regression.' (*ibid*, p.51). A larger sample from the onset as well as a more careful assignment of identification codes to our subjects would have prevented these flaws, which is an important lesson for future impact evaluation work. Yet, regardless of these limitations, we expect that a larger sample size would not lead to changing signs of the relationships we found.

Two final remarks should be made. First, the 'community-driven' elements in the project under evaluation are limited, and so are our results. It would be worthwhile to study whether more intensive involvement of project participants and more salient mechanisms of guaranteeing the quality of group leadership would benefit the project impact, using a larger sample of communities. Second, our data do not allow us to measure eventual productivity differences between private and communal farms, and related allocative efficiencies of production factors. This remains a key avenue for future research.

References

Angrist, J.D. & Pischke, J.-S., 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton: Princeton University Press.

Arimond, M. & Ruel, M.T., 2004. Dietary diversity is associated with child nutritional status: evidence from 11 demographic and health surveys. *The Journal of Nutrition*, 134(10), pp.2579–85.

Bauer, M. et al., 2014. War's enduring effects on the development of egalitarian motivations and in-group biases. *Psychological science*, 25(1), pp.47–57.

Beath, A., Christia, F. & Enikolopov, R., 2012. *Power to the People? Experimental Evidence on Direct Democracy from Afghanistan*, Cambridge, MA.

Beekman, G., Bulte, E.H. & Nillesen, E., 2013. Corruption and economic activity: Micro level evidence from rural Liberia. *European Journal of Political Economy*, 30, pp.70–79.

Van den Berg, H. et al., 2007. Investing in Farmers: The Impacts of Farmer Field Schools in Relation to Integrated Pest Management. *World Development*, 35(4), pp.663–686.

Bernard, T., de Janvry, A. & Sadoulet, E., 2010. When Does Community Conservatism Constrain Village Organizations? *Economic Development and Cultural Change*, 58(4), pp.609–641.

Björkman, M. & Svensson, J., 2009. Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monitoring in Uganda. *The Quarterly Journal of Economics*, 124(2).

Blattman, C. & Miguel, E., 2010. Civil War. *Journal of Economic Literature*, 48(1), pp.3–57.

Burde, D., 2004. Weak State, Strong Community? Promoting Community Participation in Post-Conflict Countries. *Current Issues in Comparative Education*, 6(2), pp.73–87.

Casaburi, L., Glennerster, R. & Suri, T., 2013. *Rural Roads and Intermediated Trade: Regression Discontinuity Evidence from Sierra Leone*, Cambridge, MA.

Casey, K., Glennerster, R. & Miguel, E., 2012. Reshaping institutions: Evidence on aid impacts using a preanalysis plan. *Quarterly Journal of Economics*, pp.1755–1812.

Colletta, N. & Cullen, M., 2000. *Violent Conflict and the Transformation of Social Capital: Lessons from Cambodia, Rwanda, Guatemala, and Somalia.*, Washington, DC: The World Bank.

Collier, P. et al., 2003. *Breaking the Conflict Trap: Civil War and Development Policy*, Washington, DC: World Bank.

Collier, P., 1999. On the economic consequences of civil war. *Oxford Economic Papers*, 51(1), pp.168–183.

Davis, K. et al., 2012. Impact of Farmer Field Schools on Agricultural Productivity and Poverty in East Africa. *World Development*, 40(2), pp.402–413.

Ellis, F., 1993. *Peasant economics: Farm households in agrarian development*, Cambridge, UK: Cambridge University Press.

Ellis, F., 2006. *The Mask of Anarchy*, New York: New York University Press.

Fearon, J.D., Humphreys, M. & Weinstein, J.M., 2009. Can Development Aid Contribute to Social Cohesion after Civil War? Evidence from a Field Experiment in Post-Conflict Liberia. *American Economic Review*, 99(2), pp.287–291.

Fearon, J.D., Humphreys, M. & Weinstein, J.M., 2013. *How Does Development Assistance Affect Collective Action Capacity? Results from a Field Experiment in Post-Conflict Liberia*, Stanford, CA.

Friis-Hansen, E. & Duveskog, D., 2012. The Empowerment Route to Well-being : An Analysis of Farmer Field Schools in East Africa. *World Development*, 40(2), pp.414–427.

Gilligan, M.J., Pasquale, B.J. & Samii, C.D., 2010. *Civil War and Social Capital: Behavioral-Game Evidence from Nepal*,

Gneezy, A. & Fessler, D.M.T., 2012. Conflict, Sticks and Carrots: War Increases Prosocial Punishments and Rewards. *Proceedings of the Royal Society B: Biological Sciences*, 279, pp.219–223.

Grossman, G. & Hanlon, W.W., 2014. Do Better Monitoring Institutions Increase Leadership Quality in Community Organizations? Evidence from Uganda. *American Journal of Political Science*, 58(3), pp.669–686.

Guo, M. et al., 2015. Farmer field school and farmer knowledge acquisition in rice production: Experimental evaluation in China. *Agriculture, Ecosystems and Environment*, 209, pp.100–107.

Humphreys, M., Sanchez de la Sierra, R. & Windt, P. Van Der, 2015. *Social Engineering in the Tropics: A Grassroots Democratization Experiment in the Congo*, Cambridge, MA.

IMF, 2008. *Liberia: Poverty Reduction Strategy Paper*, Washington D.C.

Jones, B.F. & Olken, B.A., 2005. Do Leaders Matter? National Leadership and Growth since World War II. *Quarterly Journal of Economics*, 120(3), pp.835–864.

Khwaja, A.I., 2009. Can good projects succeed in bad communities? *Journal of Public Economics*, 93, pp.899–916.

King, E., 2013. Can Development Interventions Help Post-conflict Communities Build Social Cohesion? The Case of the Liberia Millennium Villages. *CIGI-Africa Initiative Discussion Paper Series*, (9).

King, E. & Samii, C., 2014. Fast-Track Institution Building in Conflict-Affected Countries? Insights from Recent Field Experiments. *World Development*, 64, pp.740–754.

Kyamusugulwa, P.M. & Hilhorst, D., 2015. Power Holders and Social Dynamics of Participatory Development and Reconstruction: Cases from the Democratic Republic of Congo. *World Development*, 70, pp.249–259.

Larsen, A.F. & Lilleør, H.B., 2014. Beyond the Field: The Impact of Farmer Field Schools on Food Security and Poverty Alleviation. *World Development*, 64, pp.843–859.

Mansuri, G. & Rao, V., 2013. *Localizing Development. Does Participation Work?*, Washington, DC: The World Bank.

Matsusaka, J.G., 2005. Direct Democracy Works. *Journal of Economic Perspectives*, 19(2), pp.185–206.

Olken, B.A., 2010. Direct democracy and local public goods: Evidence from a field experiment in indonesia. *American Political Science Review*, 104(2), pp.243–267.

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

Platteau, J.-P., 2004. Monitoring Elite Capture in Community-Driven Development. *Development and Change*, 35(2), pp.223–246.

Platteau, J.P. & Gaspart, F., 2003. The risk of resource misappropriation in community-driven development. *World Development*, 31(10), pp.1687–1703.

Porter, G., 2002. Living in a Walking World: Rural Mobility and Social Equity Issues in Sub-Saharan Africa. *World Development*, 30(2), pp.285–300.

Reno, W., 1995. Reinvention of an African patrimonial state: Charles Taylor's Liberia. *Third World Quarterly*, 16(1), pp.109–120.

Richards, P. et al., 2005. *Community Cohesion in Liberia: A Post-War Rapid Social Assessment*, Washington DC: World Bank.

Rosenbaum, P.R. & Rubin, D.B., 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), pp.41–55.

Swindale, A., Bilinsky, P. & Swindale, A., 2006. *Household Dietary Diversity Score (HDDS) for Measurement of Household Food Access: Indicator Guide (v.2)*, Washington D.C.

Todo, Y. & Takahashi, R.Y.O., 2013. Impact of Farmer Field Schools on Agricultural Income and Skills: Evidence from an Aid-Funded Project in Rural Ethiopia. *Journal of International Development*, 25, pp.362–381.

Voors, M.J. et al., 2012. Violent Conflict and Behavior: A Field Experiment in Burundi. *American Economic Review*, 102(2), pp.941–964.

Wairimu, W.W., Hilhorst, T. & Slingerland, M., 2014. *Aid under contestation: Everyday practise and "farmer" responses in "community-based" food security interventions in Northern Uganda*. Wageningen: Wageningen University.

APPENDIX

#	Module topic
1	Mobilisation meeting
2	Self-reliance
3	Gender in traditional Liberian society
4	Sustainable group formation / Constitution and by-laws
5	Leadership structure for effective working group/ record keeping
6	Planning the Cropping calendar
7	Site selection and soil fertility
8	Farm Management
9	Brushing and clearing
10	Seed and nursery preparation
11	Swamp layout / upland rice growing
12	Organic farming methods (composting, organic pesticides production, etc.
13	Harvesting and storing
14	Recap of previous modules
15	Quality food for better health
16	Marketing