SKILL VERSUS VOICE IN LOCAL DEVELOPMENT

Katherine Casey*	Rachel Glennerster	Edward Miguel	Maarten Voors
Stanford Graduate School of Business and NBER	Department for International Development	University of California, Berkeley and NBER	Wageningen University
kecasey@stanford.edu	rglennerster@gmail.com	emiguel@berkeley.edu	maarten.voors@wur.nl

THIS VERSION: 25 September 2019

Abstract

Where the state is weak, traditional authorities control the local provision of land, justice, and public goods. These authorities are criticized for ruling undemocratically, and are poorly educated relative to younger cohorts. We experimentally evaluate two solutions to these problems in Sierra Leone: one to foster citizen participation in governance and development projects; and another to identify skilled technocrats and delegate project management to them. In a real-world infrastructure grants competition, we find that a public nudge to delegate dominates both the default of chiefly control and the participatory reform. Results uncover a broader failure of autocratic institutions to fully exploit human capital.

JEL Codes: H41, I25, O15
Keywords: Governance, human capital, traditional authority, foreign aid, community driven development

---

* Corresponding author: 655 Knight Way, Stanford CA 94305, +1 (650) 725-2167.

We thank Angélica Eguiguren, Erin Iyigun, Mirella Schrijvers, Eleanor Wiseman and the Innovations for Poverty Action team in Freetown for excellent research assistance and fieldwork. We thank the Decentralization Secretariat, the GoBifo Project, Local Councillors in Bombali and Bonthe districts, and a panel of experts for their collaboration. Mike Callen, Macartan Humphreys, Ken Opalo, Ann Swidler, Eva Vivalt and numerous seminar participants provided valuable comments. We gratefully acknowledge financial support from the UK Economic and Social Research Council, the Governance Initiative at J-PAL, NWO 451-14-001 and the Stanford Institute for Innovation in Developing Economies. All errors are our own. This study was pre-registered on the AEA registry: https://www.socialscienceregistry.org/trials/1784.

I. Introduction

Rural communities in poor countries often fall beyond the reach of the formal central state and must provide a variety of essential public goods and services for themselves. Such provision requires fundraising external capital, usually from other government agencies or non-governmental organizations, and then managing technical aspects of project implementation. The traditional authorities who typically govern this process are criticized for falling short in two ways: first, they often rule in a largely undemocratic and unaccountable fashion; and second, they tend to be quite old, and given the recent investments in education in poor countries, they are often far less educated than younger community members. This naturally raises the question of whether increasing citizen voice in governance, or better leveraging existing human capital and skill, could be effective responses to the challenges of local development.

These are major concerns in Sierra Leone, our empirical setting, which sits squarely at the bottom of international rankings of government effectiveness, public services, and economic development.[1] Public goods provision, land distribution and local justice decisions are dominated by traditional chiefs who are unfettered by institutional checks and balances and face no direct electoral pressure. There is evidence that the more politically powerful these chiefs are, the worse are long-run development outcomes (Acemoglu, Reed and Robinson 2014). Yet the present is also a time of rapid societal change and opportunity: after decades of profound neglect—fully 71% of Sierra Leoneans in 1985 had never been to school—the government and its donor partners have achieved universal primary enrollment since the end of the country's civil war (1991-2002).[2] We explore how traditional authorities respond to this sharp increase in the human capital stock: do they harness these skills for the more technical aspects of development, or do they sideline the new talent, who are by definition not part of the elder ruling elite and thus a potential political threat?

These two limitations facing traditional authorities relate to a lively debate in macroeconomics about the "deep" determinants of economic growth. Some scholars argue for the importance of institutional constraints on executive power in facilitating growth (North 1990, Acemoglu, Johnson and Robinson 2001, Rodrik et al 2004), which seems relevant given how

[1] For example, Sierra Leone is in the 10th percentile of government effectiveness according to the World Bank (http://info.worldbank.org/governance/wgi/#reports), and ranks 179th out of 188 in the United Nation's Human Development Index (http://hdr.undp.org/sites/default/files/rankings.pdf).

[2] Central Statistics Office (1985) for the educational attainment among those five years and older in 1985, and World Bank (2017) for current school enrollment rates.

weakly checked and widely ranging chiefly authority is in Sierra Leone and elsewhere (Baldwin 2016, Bulte, Richards and Voors 2018). Others counter with evidence for the importance of human capital for economic growth (Barro 1991) and its possible primacy over institutions as a driver of development (Glaeser et al. 2004, Gennaioli et al. 2013). Here we are interested in whether communities are effectively harnessing their newly enhanced stock of human capital. Yet policymakers implement programs and not abstract ideas, so face the challenge of distilling these arguments into feasible, actionable interventions. This paper experimentally examines two such policy translations, evaluating not the big ideas themselves, but rather whether policy attempts to either increase voice or to leverage skill work better in unlocking development opportunities.

Regarding the first, the international aid community operationalizes the idea of introducing executive constraints via promoting popular participation: give citizens more voice in development programs under the assumption that the input and oversight they provide will curtail the power of the elite (White 1999, Mansuri and Rao 2013). We study a commonly deployed version of this approach, called community driven development (CDD), which provides funding for local public goods construction and requires communities to make planning and implementation decisions in an inclusive and democratic manner. The World Bank, for one, dedicates 5 to 10 percent of its global portfolio to CDD projects, with over $17 billion in active investments.[3] Within this type of aid, the specific project considered in Sierra Leone represents an upper bound on the intensity of resources dedicated to facilitating broad-based participation (Casey 2018, pg. 145).

In earlier work, we evaluated the short-run effects of this program and found little evidence that it transformed local institutions (Casey, Glennerster and Miguel 2012), so why should we expect anything to be different now? First, institutions evolve slowly, so early seeds of change may need time to reach fruition (Woolcock 2013), and we have expanded the time horizon to over a decade. Aligned with this view, we elicited prior beliefs from experts in both policy and academia about the prospects for long-run change, and they on average were optimistic: they forecasted a sizeable positive effect on institutions (in 2016) that is statistically larger than what we documented earlier (in 2009). Second, these communities received a subsequent round of assistance after our earlier data collection, so we can now measure the effects of a larger "dose" of CDD. And third, we will implement a more proximate test that focuses on participation as an intermediate outcome, which could itself help unlock development opportunities, even if it does not impact the ultimate

[3] Independent Evaluation Group (2017): https://ieg.worldbankgroup.org/Data/reports/lp_genderincdd_01272017.pdf

outcome of democratizing traditional institutions. On this point, Anderson and Magruder (2017) reanalyze our 2009 data using more flexible econometric methods and find evidence for robustly significant short-run effects of CDD specifically on measures of participation in local governance.

The other intervention we study focuses on leveraging human capital by encouraging communities to delegate technical tasks to those best able to complete them. In particular, we study a low-cost, two-pronged approach to improve the skill level of managers in charge of local development projects. The first component uses a combination of community nominations and objective written tests to identify high skill local residents and delegates the sourcing and implementation of public infrastructure projects to them; the second component provides practical training to these "technocrats" in the nuts and bolts of project management.

This focus on delegation to technocrats relates to long-standing arguments about the importance of state capacity and the competence of public sector workers (Huntington 1968), which could be particularly impactful in poor countries (Finan, Olken and Pande 2017). The emphasis on delegation is motivated by the theoretical insights of Alesina and Tabellini (2007, 2008) who identify conditions under which it may be optimal to allocate tasks away from politicians and instead give them to bureaucrats.[4] The technical nature of many aspects of development projects—including infrastructure costing, contracting and engineering—combined with the relatively low skill level of chiefs, makes their management a prime candidate for delegation in settings like Sierra Leone. Technical demands also suggest that practical training in project management might be useful in further facilitating local development.

We evaluate these two distinct approaches—delegation versus broad participation—and compare them to the default of chiefly dominance, in the context of a real-world grants competition run by the district government. We find that the low-cost test quickly identified community members with significantly stronger project management skills than local chiefs. In a main finding, we show that putting these technocrats in charge of the community's application for the grants competition dominates both the default of chiefly control and the long-run CDD program. In particular, we find large positive effects of technocratic selection on objective measures of proposal quality, as well as the likelihood of being awarded an infrastructure grant. Offering training to these high skill individuals generates additional gains in performance. In contrast,

[4] A similar tradeoff between reliance on bureaucratic capacity and affording citizens greater voice has been shown to be important for the introduction, and subsequent undermining, of Chinese local elections (Martinez-Bravo et al 2017).

outcomes for the CDD communities are statistically indistinguishable from the controls, despite the closeness of our test to the activities undertaken as part of the CDD project and the intervention costing an order of magnitude more than the technocratic selection approach.

We then explore the mechanisms that appear to explain why technocratic selection and training were successful and why the emphasis on participation largely failed in these regards.

First, we find that when left to their own devices, chiefs fail to delegate complex project tasks to high skill community members, even when it appears to be in the community's interest to do so. Our setting provides a particularly stark illustration of this misallocation of human capital, given that basic literacy and numeracy are clearly valuable in drafting a successful proposal, which many chiefly elites do not possess; and in light of the considerable financial resources that high quality proposals could unlock in the government's infrastructure competition. This establishes an empirical proof of failure of the "political Coase theorem:" while delegation would clearly increase local output, some of which the chiefs could redistribute to themselves, they instead make the inefficient and costly choice to retain control of the process (Acemoglu 2003).

This has implications beyond this specific experiment, as it suggests that traditional authorities are not optimally adapting to the large positive shock to human capital that has occurred in recent decades. The skills of younger, more educated cohorts are thus considerably underutilized in the prevailing approach to local development. While this project focuses on the fairly narrow (though critical) task of securing external funding, the general point may hold for a range of other local governance tasks that rely on technical skill, including budgeting and planning, tax collection, and interfacing with the formal state as it decentralizes (a process launched in 2004), all of which fall under the purview of traditional rural authorities.

Second, we find that the light touch selection intervention may help to correct this failure to harness local skill. Specifically, technocratic selection worked in this environment because community members—including some, though not all, chiefs—on net responded positively to objective information about which local individuals were high skill, and were willing to delegate project management to them when publicly encouraged to do so. An immediate concern is that technocrats, even with supplemental training, may falter at project implementation since they lack the chief's political authority and experience. Counter to this view, however, data from physical assessments of all infrastructure built through the grants competition reveal no statistically significant differences in the quality of projects managed by technocrats versus chiefs.

Third, CDD communities were, by contrast, largely unsuccessful at bringing high skill individuals into local public service. While we do find that chiefs in CDD villages are somewhat more likely to delegate to a high skill resident than in the controls, consistent with CDD enhancing participation, this small shift does not meaningfully affect performance in the grants competition. Along several additional measures beyond the government grants competition, including the use of democratic processes in public deliberation and the local response to the 2014 Ebola public health crisis, CDD communities do not appear demonstrably more inclusive or effectively governed relative to control villages.

This is not to say that the overall CDD package of support, which includes both project funding and efforts to promote participation, is without value. Indeed, we find evidence for persistent gains in local public goods and market activity in CDD communities over 10 years after the program launched, albeit with some decay over time. Our ability to assess long-run effects of CDD is a novel contribution of this study, and is especially important in light of CDD's prominence in debates about the effectiveness of foreign aid. In particular, given the challenges of the post-conflict operating environment, the durability of the public goods constructed with CDD funds and their associated material benefits is notable. We compare our estimates about the long-run effects of CDD, as well as the relative efficacy of technocratic selection versus CDD, to the expert forecasts that we elicited before analyzing the data (following DellaVigna and Pope 2018).

The rest of the paper is structured as follows. Section II describes the experimental design and details of the two interventions. Section III, IV and V present the main empirical findings and mechanisms relating to delegation and performance in the infrastructure grants competition. Section VI explores broader impacts of the CDD program on public goods and democratic practices in the long run. Section VII assesses the extent to which the results align with the prior beliefs of experts. The final section concludes with discussion of policy implications and the generalizability of results both within Sierra Leone and beyond.

II. Experimental Design and Interventions

This research was designed around a real-world economic development opportunity. In 2016, the elected district governments (called Local Councils) in our Sierra Leone study areas ran a competition to award grants for small-scale infrastructure construction. Entering the competition required a detailed project proposal and budget (three pages in length), submitted to the district

government office. A committee of elected Local Councillors evaluated and ranked all proposals, blinded to the name of the submitting village, and awarded implementation grants each worth $2,500 to the top twenty proposals. Communities were informed about the number of winning proposals, but not about the pool of eligible villages nor about the likelihood that other communities would apply (which we ourselves could not know *ex ante*), which suggests there was considerable uncertainty about the odds of success. A $2,500 grant is sufficient to fund the construction of a community center, grain storage house or multiple latrines in one of these communities, which are meaningful projects. As we show later, 98% of villages entered the grants competition, which provides revealed preference evidence that communities found this a worthwhile opportunity.

Local Councils are relatively new in Sierra Leone, created by the Local Government Act of 2004. Prior to that, these rural communities were relatively untouched by the modern central state—which has been characterized as highly corrupt, incompetent and authoritarian (Reno 1995)—and instead governed largely by traditional authorities. At the community level, the village headman, who occupies the most local tier of the chiefly hierarchy, remains the most influential leader over matters of land, labor and justice. Some scholars claim that the chiefs' exclusive leadership style, combined with vulnerability to coerced labor and capricious fines, was a key driver for young men to take up arms during the country's civil war (Richards 1996). More recent evidence suggests that the least constrained chiefs perform worse on local development, while simultaneously enjoying greater legitimacy, a combination that Acemoglu et al. (2014) interpret as evidence that "more dominant chiefs have been better able to mold civil society and institutions of civic participation in their villages for their own benefit and continued dominance" (page 323). This suggests that it might be difficult for community members, particularly those who are not part of the ruling elite, to assert themselves in matters of local development (either by participation or delegation), including the district government grants competition.

We evaluate the effectiveness of two distinct interventions in allowing communities to avail themselves of the grants opportunity, and benchmark both against the default of traditional chiefly authority. We used a cross-randomized design that overlaid a new technocratic selection intervention over the sampling frame of a long-term CDD experiment, and tracked how all communities performed in the grants competition (see Figure 1 for a description of the design).

IIA. Intervention 1: Technocratic Selection

To motivate technocratic selection, it is worth first considering the many reasons why a traditional chief might not be the best person to manage the community's entry into the district government grants competition. Alesina and Tabellini (2007) argue that it is socially optimal to delegate tasks to independent bureaucrats instead of elected politicians if the task is difficult, politician capability to execute is uncertain, or monitoring performance requires expertise. By these metrics, the grants competition would seem to sit squarely in the bureaucrat's purview. Developing a detailed proposal is technically demanding, involving planning, writing text, and budgeting. It is unclear that the traditional village headman, as the top local politician, has the requisite skills to complete it. Moreover, given the historic lack of educational opportunities in Sierra Leone, it will be difficult for most adults in the village to assess the quality of the proposal generated. We thus examine whether there are other community members, outside the chiefly elite, whose skills might be a better match for this task but are currently underutilized.

Successfully implementing a technocratic selection process requires at least two things: first, identifying individuals with the appropriate skills; and second, encouraging community leaders to delegate project management authority to them. We then complement selection with management training for a subset of technocrats (discussed at the end of this section).

To identify potential technocrats, we used a combination of community nominations and objective tests. Specifically, our field team supervisors convened a public meeting of local leaders and residents in all study communities, focused on publicizing the grants competition. The team supervisor explained the size of the grants, how the competition worked, and encouraged communities to enter. Then he (or she) went through the standardized application form and explained what was required in each section, emphasizing the skills needed to develop a successful submission, and asked the group to think of people in their community who had the appropriate skills. As an example, when the supervisor explained the budget template, he asked the group to think of people who are good with numbers and have experience costing project inputs like cement and iron sheets. Other skills emphasized include writing a persuasive project plan, time management, and the ability to get things done (see Appendix A for the implementation script). The supervisor then asked the gathered community members to deliberate and nominate five individuals, other than the local chief, who possessed the requisite skills, and the supervisor then stepped aside to allow the community to generate their list of nominees.

To complement this local knowledge, we asked all five nominees as well as the village headman to then take an objective written test in private. We designed the test to capture the skills associated with managerial capital, which scholars have found to be important for the profitability of firms in India (Bloom et al. 2013, 2018), performance of public agencies in Nigeria (Rasul and Rogger 2016), and implementation of NGO-sponsored projects in Sierra Leone (Voors et al. 2018). The test included questions that measure basic literacy and numeracy; experience implementing development projects; ability to cost a standard infrastructure project (specifically a 10 foot by 10 foot cement floor for drying agricultural goods, a common project in rural areas); and past community leadership roles. The test runs to 121 points and generated wide dispersion in scores: the range across all test-takers was 1 to 108, with a mean of 42 and standard deviation of 26. Field enumerators scored the tests on site and the highest score amongst the five (non-chief) nominees was designated as the local technocrat in the treatment communities (discussed further below).

The second component of technocratic selection is delegation of project authority, which we exogenously varied across communities via a public "nudge." After scoring the managerial capital tests, the field supervisor reconvened the community meeting. He explained that he would unlock a lottery which would determine whether the person with the highest score (of the five non-chief nominees) should be put in charge of managing the project challenge submission, or whether the community should rely on the chief as usual. The supervisor then held up a tablet device with a rolling dice visual lottery image that broke apart into the assignment screen, which read either "Highest scorer" or "Headman leader." The nudge towards delegation to the highest scorer is our technocratic selection treatment, while the reversion to the chief as usual is the default condition. Neither condition publicly announced any of the individual test scores. The supervisor then made a display of writing the assigned person's name at the top of the project challenge application, handing the application over to that person, and giving him or her a voucher to subsidize their transport to deliver the community's submission to the relevant district government office.

Note that while the announced nudge was public, there is nothing binding about the encouragement to delegate to the technocrat. There are, moreover, several reasons to believe that a nudge to delegate would have little effect on the nature of project development. First, the communities were informed that the grants competition was run entirely by the local government (and not the research team, see Appendix A for the supervisor script), so there was no obvious need to comply with the suggested delegation nudge. Second, if traditional authorities recognize

that technical skills matter for project success, and they have good information about local citizens, chiefs may already be delegating project management efficiently in the status quo. In other words, if chiefs know which local residents can read and write and have project experience, they may willingly choose to delegate complex tasks to these high skill individuals. And moreover, if they learn something about local skills from the community nomination process, they might become more willing to delegate even in the default condition, suggesting our estimates would represent a lower bound on the impact of technocratic selection.

On the other hand, technically competent managers might lack the authority or political influence of traditional leaders, leading them to fail at project management. For instance, the younger cohorts who benefited from educational expansion and the teachers hired to staff local schools may not be able to mobilize labor and financial contributions from other community members as effectively as chiefs, or even determine which project is needed. This could lead communities to choose chiefly authorities to manage the project regardless of the nudge they received from the research team. Or, if the traditional chiefs see these high human capital managers as a political threat, they may try to sideline them from the process or sabotage their efforts. Any combination of these factors would work against finding a treatment effect of the technocratic selection nudge on performance in the grants competition.

One final concern is that the selected technocrats, while possessing greater general human capital, may not yet have the specific skills needed to write a strong grant proposal or manage a public project. This suggests that training could be valuable in this setting, and particularly so if they are fast learners, i.e., there is complementarity between the training and underlying human capital. Outside of frontline service providers (for instance, teacher training), there is limited rigorous evidence on the effectiveness of public sector management training in low-income countries (Finan, Olken, and Pande 2017). There is some evidence that managerial practices can be effectively taught in formal private sector firms (Bloom et al. 2013), but results for training small-scale entrepreneurs are more pessimistic (see McKenzie and Woodruff 2014).

To examine the impact of training individuals, we subsidized the cost of attending one of several all-day, small-group courses focused on basic project management skills. The courses covered budgeting, accounting, planning, and grant writing, and were run by the respective ward development committees (the head of which is an elected member of district government) in partnership with a local consultant, as part of the broader grants competition. To evaluate their

efficacy, the research team offered an attendance subsidy to a randomly chosen half of the selected technocrats (no subsidy was offered to chiefs in the default condition, see Figure 1). In these subsidy communities, the field supervisor concluded the community meeting by providing the date and location of the nearest training, informed the community that the travel costs of the selected manager would be reimbursed, and encouraged that person to attend the training.

IIB. Intervention 2: Community Driven Development

The technocratic selection arms cross the experimental frame of an existing long-run community driven development (CDD) study, see Figure 1. The CDD project, called GoBifo (which means "move forward" in the local Krio language), was funded by the Government of Sierra Leone and the World Bank, and comprised of two main elements: block grants provided to communities to fund public infrastructure; and intensive social facilitation to promote broad-based participation in local governance and development programming. Project activities began by establishing a village development committee (VDC), mandated to include representatives of marginalized groups, which was trained and encouraged to make the selection, planning and implementation of community projects in an inclusive and democratic manner. The VDC was then given an opportunity to learn-by-doing in managing a series of small-scale public projects funded by the grants. We test whether the chiefs and other community leaders that have thereby been encouraged over several years to manage development projects in a more participatory way are more likely to delegate, or otherwise better leverage local talent, in the new infrastructure grants competition.

The first intense phase of GoBifo project implementation ran from 2005 to 2009 and included roughly $5,000 dollars in block grants per community (amounting to approximately $100 per household) for the construction of small scale infrastructure (like latrines, midwife huts, grain drying floors), agricultural inputs, and small business training and start-up capital. GoBifo also provided six months of dedicated organizing in each community (spread out over these first 3.5 years) to establish new institutional structures to facilitate collective action (i.e., the VDC) and put in place participation requirements to elevate historically marginalized groups—most notably women and young men—to positions of authority. The facilitation component was relatively expensive: facilitation costs 63 cents for every dollar provided in block grants, and reaches roughly one-to-one in spending if program overhead and administration are considered. To formally link project activities to higher tiers of government, the VDCs were required to submit their village

development plans to the appropriate ward development committee for review, endorsement and onward transmission to the elected district councils for approval (GoBifo Project 2007).

A second less intensive phase of GoBifo began in 2010 with additional grant support to 60 of the 118 treatment communities. These communities each received $1,300 to support youth empowerment activities ("youth" is defined by the government as individuals under 35 years of age); once again, no activities were implemented in the GoBifo control communities. Facilitation staff in both district headquarters (as well as management staff in the capital) were employed full time throughout this second period, and remain on government payroll at the time of writing. They have continued some project facilitation activities in treatment villages, although we lack reliable data on the frequency of these interactions, and our impression is that the level of support for treatment villages was minimal post-2012.

Total project costs for the first phase (2005-2009) are approximately $2 million, and for the second, less active phase (2010-present) are nearly $3 million, given the continuation of project staffing, transport and overhead for several years. The relatively high cost of the social facilitation component of CDD serves as further motivation for the technocratic selection intervention, which is far less expensive and more immediate. From the perspective of CDD treatment communities, this analysis evaluates the long-run persistence of direct programming support that largely concluded by 2012. From a broader policy perspective, we evaluate a $5 million investment in CDD that has been at least nominally operational from 2005 to the time of writing.

In data collected in 2009, shortly after the intense first phase of project activity concluded, we found evidence for substantial positive effects of these investments on the stock and quality of local public goods, accompanied by improvements in material welfare, as captured by household assets and market activity (Casey, Glennerster and Miguel 2012). At that time, we found no evidence of CDD impacts on a rich set of measures designed to capture institutional change and social capital. Short-run results from other large scale experiments in Afghanistan (Beath, Christia and Enikolopov 2013), the Democratic Republic of Congo (Humphreys, Sanchez de la Sierra and van der Windt 2015) and Liberia (Fearon, Humphreys and Weinstein 2015) are similarly mixed, however together provide little support for institutional transformation. If we narrow consideration to more modest claims about boosting participation in local governance specifically, Anderson and Magruder (2017) reanalyze our 2009 data and show that this is the one hypothesis, under our broader "family" of 9 institution-related hypotheses, that we would have found robust statistical

support for had we used more flexible, and thus higher powered, econometric methods.

The null result on institutions writ large led to some criticism that the 3.5 year evaluation timeline may have been too short to capture impacts on slowly evolving institutions, especially if institutional change follows a non-linear trajectory (Woolcock 2013). Partially in response to this perspective, we designed the data collection featured in this paper, where we returned (in 2016) to all 236 originally sampled communities, seven years after the short-run data collection (in 2009), in order to assess long-run changes in institutions, as well as assess the persistence of CDD investments in local public goods. Data collection on these long-run CDD outcomes occurred earlier in the same day that the community meeting and technocratic selection nudge were administered. Note further that the 2016 survey round incorporates the additional "dose" of CDD funding and programming that began implementation in 2010.

Returning to the overall research design in Figure 1, the crossed experiment allows us to evaluate the pure performance effect of technocratic selection in the district government grants competition (arm 2) in comparison to that of autocratic chiefs in the default condition (arm 1), and to chiefs who have been encouraged to govern more inclusively through several years of CDD programming (arm 4). It also gauges the efficacy of basic management training for high skill community members (arm 3), and captures potential interaction effects between technocratic selection, training and CDD (in arms 5 and 6).

III. Main Empirical Results

We first examine the impacts of technocratic selection and CDD on community performance in the project challenge grants competition, estimating the following model:

$$Y_c = \beta_0 + \beta_1 TS_c + \beta_2 CDD_c + \beta_3 TS_c * CDD_c + W'_c \Psi + X'_c \Gamma + \varepsilon_c \quad (1)$$

where outcome *Y* (i.e., proposal quality, winning a grant) is measured for each community *c*; *TS* is an indicator variable equal to one for assignment to technocratic selection (with or without training) and zero otherwise; *CDD* is an indicator for participation in the long-run GoBifo program; W_c is a vector of stratification fixed effects for geographic wards; X_c are balancing variables used in the original CDD randomization (community size and distance to nearest road); and ε_c is an idiosyncratic error term. The first tests of interest compare technocratic selection and CDD, respectively, to the default of chiefly dominance ($\beta_1 = 0, \beta_2 = 0$). The next test captures

the relative efficacy of technocratic selection versus CDD ($\beta_1 = \beta_2$). We also test for interaction effects between the two interventions ($\beta_3 = 0$), noting that we are somewhat underpowered statistically for this test unless effects are quite large. All estimates are intention-to-treat effects. Appendix F includes our pre-analysis plan with annotation that links each specification therein to the relevant table in the main text and appendices.

Outcomes of interest include three distinct measures of proposal quality, all based on blinded review by different sets of local development professionals in Sierra Leone, and the probability of winning an implementation grant. The first quality assessment, labeled "technical score" in Table 1, is a simple coding of proposal completeness. Local research assistants rated several binary indicators of whether the submission includes items specified in the application form (e.g., if the instructions for project description ask for four items, does the proposal contain all four?). The second, "expert score," was completed by two Sierra Leonean development practitioners not affiliated with the GoBifo project or the district governments. These experts comprehensively scored the quality of the submission with reference to the scoring guidelines used by the district governments. Third, we have the official scores for all proposal submissions and grant award decisions made by the district governments themselves. Note that we do not examine effects on entry into the competition as we originally intended, as nearly all study villages (232 out of 236) submitted a proposal, affording minimal variation to examine.[5]

Table 1, Panel A reports the first set of results. Estimates in the first column compile the three different expert evaluations into a single equally weighted index. The treatment effect estimate is 0.397 standard deviation units (standard error 0.164) for technocratic selection, indicating that communities nudged to delegate to a high skill manager submitted proposals of substantially higher quality than those in the default condition of chiefly control (that did not participate in CDD). Estimates for each of the three distinct quality assessments are all positive in sign and two are significant at 95% confidence.[6] Estimates in column 5 suggest that technocratic selection increased the probability of winning an implementation grant by 10 percentage points, a large and highly significant effect, as compared to traditional chiefly dominance.

The five analogous treatment effect estimates of CDD are much smaller in magnitude and

[5] Submission rates are statistically balanced across treatment arms and range from 97 to 100 percent.
[6] Missing scores for the four communities that did not submit a proposal are imputed at treatment arm mean. Appendix Tables A2 and A3 present imputation bounds that instead use the lowest (highest) observed score, which have little effect on the estimates.

none are statistically distinguishable from zero at conventional levels, indicating that in the multi-year participatory intervention did not substantially alter community ability to access a new funding opportunity. Estimates in the third row provide no evidence for significant interaction effects between technocratic selection and CDD. (For alternative specifications, see Appendix Table A4 for the fully interacted model and Table A5 for a simple two-way comparison of CDD to technical selection with no interaction terms. In Table A5, the *F*-test strongly rejects that the coefficients on the quality index are the same under technocratic selection versus CDD.)

We next separately estimate effects of management training beyond technocratic selection alone. In light of the null results for CDD above, we pool these arms across the CDD experiment to bolster statistical power and do not include interactions. We estimate the following model:

$$Y_c = \delta_0 + \delta_1 TS_c + \delta_2 TR_c + W'_c \Psi + \varepsilon_c \quad (2)$$

where variables remain as defined in Equation (1), save the new *TR* term that is an indicator for assignment to management training and captures the marginal effect of training beyond the effect of technocratic selection, and W_c, the vector of stratification fixed effects for geographic wards, is now interacted with CDD assignment (thus controlling for any CDD effects).[7]

Results are presented in Table 1, Panel B. The estimated treatment effect for technocratic selection alone is a 0.315 standard deviation units improvement in the proposal quality index (standard error 0.138), as compared to project management under the default of chiefly control. There is also a positive and significant additive effect of management training. The ITT effect of subsidizing travel to the training session increased the quality of the proposals generated by these technocrats by 0.339 standard deviation units (standard error 0.133).[8] Taking the two effects together, project proposals in villages that received the nudge for selecting the high skill individual and the subsidy for the management training course scored 0.65 standard deviation units higher than control villages, a very large and highly significant effect (the *F*-test rejects that both estimates are equal to zero at 99% confidence). This pattern of results is consistent across the various types of proposal evaluations: all six point estimates are positive and five are at least marginally significant. While the technocrats' proposals were of higher quality, this did not significantly affect whether or not communities won an implementation grant in this regression specification:

[7] This deviates from our PAP and is a correction to control for CDD assignment while estimating technocratic selection effects. As treatment assignment is balanced within these blocks, it makes little difference for the results.
[8] In total, two people assigned to the training did not show up and four people not assigned were trained.

estimates in column 5 (of Table 1, Panel B) are positive but not statistically distinct from zero (0.067 with standard error 0.044).

Focusing on the actual threshold for winning a grant estimates effects above the 90th percentile of the score distribution. As this threshold is quite competitive and somewhat arbitrary, as it is determined by the government's budget, it is informative to look for potential shifts in other parts of the score distribution. Figure 2 presents the cumulative density of government proposal evaluations for technocratic managers and chiefs, where it is clear that the distribution of technocrats' scores dominates, as it is shifted to the right over the entire distribution (a Kolmogorov-Smirnov test rejects equivalence at *p*-value = 0.03). The vertical line demarcates the score cut off that determined which proposals were actually funded. If we relax this, e.g. explore what would happen if the government had had more funds to allocate, we see that there are strong positive effects on winning a grant at other simulated thresholds, like the 50th percentile (see *F*-tests in Appendix Table A6).

Figure 3 summarizes these results by plotting the mean proposal score index for each of the six experimental arms. Scores are standardized with respect to chiefs in the default condition without CDD exposure (Arm 1 from Figure 1), where the mean score by construction is zero. The narrower bracket above the point estimates compares scores in Arm 2 to Arm 1 to capture the "pure" effect of technocratic selection in the absence of CDD. Here the positive and marginally significant difference in means suggests that technocrats outperform chiefs by 0.35 standard deviation units (where the associated *p*-value from a *t*-test of equivalence across arms rejects at 90 percent confidence). Comparing Arm 3 to Arm 1 reveals a positive and highly significant combined effect of selecting and training technocrats, who outperform chiefs by 0.50 standard deviation units (*p*-value = 0.02). By contrast, the three brackets below the point estimates do not find much evidence for a CDD effect. The first two estimate are null, suggesting that neither chiefs nor technocrats perform any better in CDD versus control communities. While the rightmost comparison suggests that the training of technocrats had a larger effect in CDD communities, the relevant interaction term in the regression counterpart of these estimates is not significant (in Appendix Table A4, which further includes the randomization strata and balancing variables). Moreover, the *F*-test at the bottom of Table A4 cannot reject that the three CDD-related coefficients are jointly equal to zero (*p*-value =0.23). By contrast, the *F*-test for the four coefficients related to technocratic selection and training rejects at above 99 percent confidence.

These differences raise the question of why communities do not do more to seek out technically competent managers to improve their chances of winning outside funds, an issue we explore below. It is also striking that the intense CDD program was not successful in encouraging appropriate delegation, despite its high cost and focus on facilitating broad participation in development programming, including for tasks not dissimilar to what was required in the government grants competition. The direct facilitation costs per community for the first intense phase of GoBifo (2005 to 2009) was $3,072, and adding project oversight and management brings this figure up to $5,325, a figure that excludes the substantial value of infrastructure grants; adding facilitator wages over the second less intense period (2009 to 2016) roughly doubles this cost. In contrast, implementing technocratic selection involves field visits and administering written tests, which cost just $231 per community, while the one day of basic management training costs $68 per participant, leading to a combined total of $299 per community in villages that received both. Thus CDD's facilitation cost alone is a full order of magnitude greater than the technocratic interventions, and took years to implement, in contrast to a few days.

While our objective in this experiment was to test whether communities allocate tasks to those best able to deliver them, we can also ask whether this particular version of technocratic selection is cost effective in its own right for this particular grant opportunity. For winning a grant at the actual threshold, the expected value of selection and training combined does not quite cover its cost (e.g. from Table 1, column 5 the expected value is 0.102*$2,500 = $255 < $299). This calculation would reach break-even for slightly larger grant awards ($2,960) or for lower winning thresholds (e.g. at the simulated 50^{th} percentile threshold, the value well exceeds the costs 0.146 *$2,500 = $365 > $299).

The primary cost comparison between the technocratic approach versus CDD warrants two important observations. First, technocratic selection is viable in part because donors and the Sierra Leone government have spent millions of dollars educating young Sierra Leoneans since the end of the civil war in 2002, creating a local pool of high skill young people and making technocratic selection look cheap. In settings where universal education has not been established, large human capital investments would be required. Second, the GoBifo CDD project may have many other benefits beyond performing well in the infrastructure grants competition, which are not considered here. In later discussion, we extend analysis to a variety of other outcomes and do find evidence for positive and persistent effects of CDD on other local development outcomes.

Section IV: What the Success of Technocratic Selection Implies

To better understand why the nudge toward technocratic selection had positive impacts, we consider links in the underlying causal chain.

First, the community nomination process together with written tests demonstrates that it is relatively straightforward to successfully identify high skill individuals even in very poor communities. Comparing technocratic selection to the default of chiefly control, the highest scoring manager nominated by the community strongly outperforms the village headman, by 1.7 standard deviation units on average (standard error 0.14), on the written management test. This large difference substantiates the hypothesis that there is a reserve of human capital located outside the traditional chiefly elite.[9]

Second, the written test scores are informative of performance in the district government grants competition. There is a positive correlation between the score of the selected project manager (whether chief or top scorer) and outcomes in the competition: a one standard deviation increase in test score improves measured proposal quality by 0.27 standard deviation units (standard error 0.05) and increases the probability of winning a grant by 5 percentage points (standard error 1.7). We can break out these correlations for each of the eight core competencies covered by the test. Of these, local infrastructure experience, literacy and numeracy have the most predictive power for proposal quality.

The high skill individuals differ substantially from traditional chiefs along observable dimensions. As presaged by the discussion of educational expansion, Table 2 shows that they are younger than chiefs (by twenty years on average), better educated (with 98 as compared to 35 percent likelihood of having some formal education), more likely to be from outside the village (by 19 percentage points), and more likely to be a teacher than a farmer. Notice that very few of the women put forward in the set of community nominees (which was one in four) came out with the highest test score, so nearly all of those identified in the technocratic selection nudge are men, and nearly all traditional chiefs are also men.

Third, a public nudge is sufficient to substantially change the likelihood that a high skill

[9] Note that we estimate a null result on whether management training further enhanced the technocrats' scores (equal to -0.027 standard deviation units with standard error 0.133), which provides a placebo test and "sanity check" on the research design, as the randomly assigned training took place after the tests were administered.

individual is put in charge of managing the community's entry into the grants competition and the subsequent project. To verify delegation in practice, we stationed field enumerators at the district government offices to survey people who submitted a proposal from any of our study communities. To allay concerns about social desirability bias, we asked for the names and local leadership positions of up to three people that were involved in specific aspects of the proposal process: who selected which type of project to apply for, developed the budget, and set the implementation timeline. We avoided any priming references to the lottery or public nudge, and matched the submitted names to the testing data *ex post*.

These survey reports about who was in charge of proposal generation differ markedly across treatment arms. In analyzing these differences, we group together reports for an array of chiefly authorities to account for the fact that chiefs have their own coterie of administrators, like the village secretary, whom they can rely upon for tasks involving literacy and numeracy (although note that results are similar when restricted to the chief only). Table 3 shows that, under technocratic selection, chiefly authorities were significantly less likely to choose the project (by 35 percentage points), write the description (by 14 points), compile the budget (by 15 points) and set the implementation timeline (by 12 points). Appendix Table A7 breaks these delegation effects out for trained versus untrained technocrats, and finds comparable results, suggesting that it is the selection nudge as opposed to training that drives delegation.

Note the presence of substantial two-sided non-compliance with the delegation nudge: in 20 percent of communities in the default condition, someone outside the traditional chiefly elite chose the project; and conversely in 45 percent of technocratic selection nudge communities, a chiefly authority still chose the project. Even so, the substantial differences in process are themselves perhaps surprising given that nothing about the public lottery and community nudge was binding: while the field supervisors explicitly encouraged communities to put the highest scorer in charge in treatment communities, there was no meaningful constraint on communities reverting to chiefly authority as soon as the research team left. If we use the compliance rates for delegating project choice to effectively capture the first stage of the intervention, this would inflate the estimated effect on proposal quality in Table 1, Panel A to a one standard deviation unit treatment-on-the-treated effect.[10]

[10] E.g., here we take the estimated coefficient and divide by the difference in treatment take up rates in treated and control arms: 0.397/(0.55-0.20) = 1.13 standard deviation units, a very large effect.

We cannot rule out that the technocratic selection intervention also relieved an information constraint regarding the existence and identity of high skill community members. Note a subtle asymmetry in our research design: while the chiefs in the default arm could always choose to delegate to one of the five community nominees, they were not informed about which of the five scored the highest on the written management test. So it could be the case that the chiefs always wished to delegate but were at an informational disadvantage in the default condition. However, this seems unlikely to fully account for the observed effects given the reported differences in who was in charge of the management process documented in Table 3. Moreover, the chief would have done fairly well by picking any one of the five nominees at random: for instance, 50% of the nominees had a test score of at least 60 points, which is twice the average score of chiefs. Even so, the information conveyed by revealing the top scorer may have been useful for hastening delegation, and since it comes at essentially zero marginal cost once the written tests are administered, seems worth retaining in any related future selection interventions.

Next consider reasons why the management training (offered to half of the technocrats) also appears to have been effective. Training materials were developed by a local expert and implemented in partnership with the district governments. They were designed explicitly to help communities develop successful submissions to the grants competition and covered topics including eliciting community needs, budgeting, and time management. We can leverage the fact that topics covered in the training curriculum do not perfectly coincide with the questions on the application form to assess the extent to which any observed training effect reflects "teaching to the test." We do not find that training created a purely mechanical "copycat" effect: trainees were not more likely to extraneously include topics in their proposals that were covered by the training but not called for on the application. At the same time, we do not find evidence that the skills taught during the training were applied to topics beyond its core curriculum: trainees were not more conscientious in how they responded to application questions on topics that were not covered by the training (see Table A8 for details). Together, these patterns suggest that the training effect is unlikely to be purely mechanical, but the extent to which the skills taught are broadly applicable beyond the grants competition is unknown.

Finally, we examine whether there a downside to technocratic selection in terms of the quality of project implementation for those communities that were awarded grants. In other words, conditional on winning, do chiefs do a better job at actually translating project funding into a

functional project, perhaps due to their local political influence and ability to marshal labor and other funding? If so, this could provide a rationale for why chiefs are often chosen for project leadership in the first place.

To assess this, field teams visited all twenty communities awarded grants in July 2018 (over a year after the grants were disbursed) to inspect the existence and construction quality of funded projects. Overall, 70% of the projects were deemed functional on the day of the visit; the mean quality score assessed by the team was 6.8 out of 10 points; communities contributed on average US$218 of their own funds on top of the grants; and 40% of projects were located near the chief's compound (Table A9). Taken together, there is no decisive evidence that project implementation is substantially better or worse under technocratic selection, as there are no statistically significant differences in these outcomes across treatment arms. While the rates of functionality, quality and contributions are higher for the default condition, note that this is based on the 4 chiefs who made it into the top 20 awards, who are likely positively selected and not representative of chiefs in general. Indeed, these 4 winners scored 22 points higher on the managerial capital test than the mean for all chiefs, an increase of 71 percent, indicative of strong positive selection.

Overall, the data indicate that high skill "technocrats" perform better than traditional authorities in taking advantage of a development funding opportunity, and they respond well to training in the nuts and bolts of management practices. There are obvious parallels between identifying the right people for these jobs and selection issues in personnel economics applied to public sector work. Besley and Ghatak (2005) argue that match quality with organizational mission can compensate for low-powered incentives, which are pervasive (where incentives even exist) in development programs. There is further evidence that higher pay attracts more competent workers to the public sector (Dal Bo, Finan and Rossi 2013), and thereby bringing in more competent teachers increases student learning (Alva et al. 2017). Even without pay differentials, the way in which jobs are advertised attracts different types of applicants who then perform differently on the job (Ashraf, Bandiera and Lee 2016, Deserranno 2017). Most closely related to our work here, He and Wang (2017) show that placing young college graduates into village government in China improves the targeting and implementation of social assistance programs. These results, together with our findings, indicate that there is substantial scope to attract high human capital individuals into local development projects to achieve positive public outcomes.

Section V: Why CDD Was Largely Ineffective in the Grants Competition

We explore the links in the same underlying causal chain outlined above to understand why the CDD experience failed to improve community performance in the grants competition, and then extend consideration (in the next section) to a panel of broader local governance and development metrics that we collected in both 2016 and 2009.

To start, chiefs in CDD communities did not conduct the initial public meeting or nomination process (to identify high skill residents) any differently than in CDD control communities. Recall that the field supervisor stepped out of the meeting to allow the community to deliberate independently over which five people possessed the specific skills needed to lead the entry into the grants competition. A team of enumerators stood outside the meeting during this process and recorded how the deliberation unfolded. By their observations, there is no discernable difference in the number of people in attendance, how many women or young men spoke up, whether the group took a vote, or whether the chief decided quickly with no input from others. On these direct measures of participation, inclusion and democratic practices, we find no meaningful effects of the multi-year GoBifo program. Across 16 related measures of the decision-making process, the estimated treatment effect is small at 0.035 standard deviation units and not statistically significant (standard error 0.044).

Second, deliberation in CDD communities did not generate a set of technocratic nominees that differ measurably on observable characteristics or test scores (Table 2, panel B). For example, the group of five nominees was no more likely to include a woman: 24 percent of nominees were women in both control and GoBifo communities. Similarly, CDD communities were no more likely to put forward younger people (if anything, they are slightly older on average), better educated people (70 versus 68 percent had been to school) or people from outside the village (20 versus 24 percent). Importantly, the nominees put forward by GoBifo communities did not perform any better on the management test: the difference in average test scores for selected managers (combining chiefs and top scorers) across GoBifo and control communities is 0.001 (standard error 0.156).[11] This further suggests that the learning-by-doing in implementing public infrastructure

[11] This is for the 236 managers (one per community) as assigned in the on-site lottery. There is also no difference for chiefs or highest scorers estimated separately, and there is no difference in the average scores of all five non-chief nominees in Table 2 Panel B.

projects over several years did not durably improve the stock of managerial capital in CDD villages, or the ability to identify people with these skills, at least as measured by this process.

Third, chiefs in CDD communities were slightly more likely to delegate project management to high skill individuals, but by less than is the case for the technocratic selection treatment group. In the full sample, chiefly authorities in CDD communities chose which project to enter into the competition 51% of the time, compared to 64% in controls (*p*-value on the difference is 0.08 in Table 3, panel B). Limiting consideration to the technocratic selection arms, chiefs were more likely to comply with the assignment to delegate project choice by 18 percentage points, which is significant at 95% confidence (panel C). Yet for the other three proposal activities (project description, budget and timeline) there are no statistically significant CDD impacts in either the full sample or in the technocratic selection subsample. There is also no evidence from textual analysis that proposals from CDD villages were any more likely to contain variants of the phrase "inclusion" that was a focus of CDD training or to reference village institutions like the VDC that had been put in place by GoBifo (see Table A10). This suggests that the CDD project's emphasis on inclusive leadership had only modest long-run impacts on local chiefs' willingness to delegate, and that the resulting reallocation of project work towards high skill community members was not sufficiently large to meaningfully affect performance in the grants competition (cf. modest improvements in leadership capital in the Liberian CDR study, Fearon et al 2015 page 467). Another explanation is that the ethos of CDD is one of broad participation rather than delegation to high skill individuals, however our experiment shows that this has consequences.[12]

One might be concerned that having previously benefited from CDD hurt a community's chances of receiving a grant, perhaps because the government prioritized communities that had not previously received assistance or because GoBifo communities, who have a stronger infrastructure stock, proposed different types of projects. We find little evidence to support either of these concerns. First recall that the government selection committee reviewed proposals with the village names redacted. To verify that this was not somehow subverted, Figure A1 plots the distribution of the government scores against the scores of unaffiliated development practitioners who used the same grading rubric, for communities under different treatment assignments. The two sets of scores are highly positively correlated (correlation coefficient of 0.87) and there is no

[12] Note the positive and statistically significant (on a naïve basis) short-run treatment effect on participation in local governance in column 4, which is a focus of Anderson and Magruder's (2017) re-analysis of our 2009 data.

apparent bias against GoBifo communities by government raters (e.g. there are not systematically more circles than triangles below the 45 degree line). What comes through clearly is that technocratic selection villages, from both GoBifo treatment and controls, score higher on both metrics (e.g. there are more shaded than hollow shapes in the upper right quadrant of the graph). Second, while GoBifo communities were marginally less likely to propose a community center (in Table A10), which was the most popular type of project funded, Table A11 shows that the main results from Table 1 are robust to including fixed effects for the type of project proposed. Table A10 also shows that technocrats did not systematically propose different types of infrastructure projects than chiefs.

Section VI: Broader Long-run Impacts of CDD

The findings above that even an intensive CDD project leads to only small improvements in delegation resonates with other local governance measures we collected. During the 2016 field visits, enumerators collected data on as many of the same indicators of institutional quality and performance that were measured in the earlier Casey et al (2012) study. Table 4, Panel A presents estimates of long-run CDD impacts on these measures. Combining all 61 individual outcomes grouped under this institutions "family" into an equally weighted index yields a positive, precisely estimated, but small in magnitude treatment effect estimate of 0.066 standard deviation units (standard error 0.025). We then break these outcomes into nine distinct hypotheses about how CDD might alter institutions, following the approach developed in partnership with the CDD practitioner team in 2005 (Casey et al. 2012). Looking at each underlying hypothesis in turn, while two are significant at conventional levels on a per-comparison basis (column 2), none are significant when adjusting for multiple inference (column 3). One way to interpret this pattern of results is that if we conceive of all outcomes measuring a latent variable associated with institutional quality and inclusion, CDD had a small positive impact, but the effect is not large enough to detect effects along any of the nine hypothesized channels.

Note that the long-run data collection focused on a subset of outcomes collected in 2009, as we did not collect data from households in 2016 due to research budget constraints, and instead focused on surveys with community leaders and direct observation of infrastructure. If we limit consideration to outcomes that were collected in identical fashion across the two survey rounds (2009 and 2016), the overall CDD treatment effect is similar for 2016 (at 0.064, with standard

error 0.027). This effect also becomes somewhat larger and is statistically significant using the same subset of outcomes from the 2009 data (0.086 standard deviation units, standard error 0.030, see Table A12). This difference could reflect differences in reporting between households and community leaders (although it is unclear to us *ex ante* which group is more or less susceptible to social desirability bias), or could simply be due to sampling variation created by focusing on a subset of outcomes. Either way, estimated treatment effects remain quite small in magnitude in both 2009 and 2016.

One way to gauge the real world import of these estimates is to focus on specific outcomes rather than indices. For example, one indicator behind the positive and marginally significant estimated effect for collective action is the presence of a village development committee. In 2016, CDD communities were 17 percentage points more likely to have a VDC, compared to 43 percent with a VDC in control villages. Yet despite having the institutional architecture in place, CDD communities were no more likely to have a village development plan: about half of all communities had one and this does not vary by CDD treatment status. This combination conjures up an image of dormant institutions that exist on paper but are not being used for much in practice.

Even so, if these latent institutions reduce the organizing cost of collective action, they could potentially be reactivated when needed. This was not the case for the new grants competition opportunity, but we are also able to test whether they yielded benefits in confronting a crisis. Specifically, we assess the effect of the long-run CDD project on community response to the Ebola epidemic, which tragically hit Sierra Leone in 2014. The crisis resulted in over 4,000 deaths in Sierra Leone alone. The two districts where GoBifo was implemented were differentially affected: Bombali saw 1,050 suspected cases and 391 deaths, while Bonthe was much less severely hit, with 5 suspected cases and 5 deaths. Some of the actions the government asked communities to take to prepare for and respond to cases—such as create community by-laws, report suspected cases and disseminate prevention information—could be facilitated by local institutional capacity of the kind GoBifo aimed to build.

In the 2016 follow-up survey, we examined a variety of outcomes that capture institutional responses, such as the creation of an Ebola task force, and knowledge about the epidemic (e.g., "how can you get Ebola?"); see Table A13 for details. Mean performance on these measures is moderate: for example, 66% of communities established a task force, and focus group participants gave the correct answer to roughly half of the questions about Ebola symptoms and prevention.

The estimated CDD treatment effect on a mean index of Ebola responsiveness is not statistically distinguishable from zero when evaluated on the full sample (coefficient 0.042, standard error 0.038) or for the harder hit Bombali district alone (0.007, 0.048, N = 156 communities). We also pre-specified a secondary analysis to examine knowledge versus action outcomes separately. For the latter, we do estimate a positive and significant effect for the two action outcomes of 0.153 standard deviation units (standard error 0.064) in the full sample, although it is not statistically distinguishable from zero in the Bombali-only subsample. Taken together, there is little evidence that the CDD program generated meaningful benefits for villages during the Ebola crisis, either.

By contrast to these null results, estimates for long-run impacts on local public goods are more favorable. Recall that Casey et al. (2012) did find large positive effects on the stock and quality of local public goods, accompanied by improvements in material welfare, equivalent to nearly 0.3 standard deviation units, as a result of financial resources transferred to CDD communities. While it may not be entirely surprising that a substantial cash infusion into poor remote communities shows up in economic measures like household assets, we would think more favorably about the CDD approach if these effects persisted over time. This is particularly true given that CDD projects tend to be implemented at lower cost than other government service delivery mechanisms (Wong 2012), raising the question of whether they were done to a lower standard. While our experimental design does not allow us to directly compare infrastructure provision under CDD versus other mechanisms, there are some useful benchmarks in the literature. Miguel and Gugerty (2005), for example, find that nearly half of borehole water wells built by a European bilateral aid donor in Kenya in the 1980's were no longer functional within a decade of construction. For this reason, CDD emphasizes the role of local participation in better aligning investments with demand and thereby bolstering utilization and maintenance over time (Dongier et al. 2002). The 2016 data we collected provides a unique opportunity to test these claims over a decade-long CDD program.

In the 2016 survey round, we again collected as many of the same infrastructure outcomes from the earlier study as possible, and organized them into three hypotheses about project implementation (e.g., does the community have a VDC?), the stock and quality of local public infrastructure (does the community have a functional water well?), and economic activity (how many goods are for sale in the community?); for details on these measures, see Appendix E.

Table 4, Panel B presents results. All three CDD treatment effect estimates are positive,

large in magnitude, and highly statistically significant (column 1), even after accounting for the fact that we are testing multiple hypotheses on the same dataset (column 3). For the overall "family" of infrastructure outcomes, the long-run effect of 0.204 standard deviation units is two thirds the size of the short run effect (shown in column 4), suggesting considerable persistence even years after most direct financial support ceased, although note that the estimated decay, where one third of the original effect has dissipated, is significant (column 5). Estimates do not change substantively when we limit the set of outcomes to those that form an exact panel (which includes 28 of the original 39 outcomes from 2009): the 2016 treatment effect estimate is 0.208 standard deviation units (s.e. 0.041) and that for 2009 is 0.352 (s.e. 0.035), see Appendix Table A12.

Examining each hypothesis individually, the largest estimated decay is apparent for project implementation. To give a sense of magnitude, the 17 percentage point greater likelihood of having a VDC in 2016 (mentioned earlier) corresponds to a 40 percentage point difference back in 2009 (the prevalence in control communities has remained roughly constant over time). By contrast, there is no statistically detectable difference from the short- to long-run for the impacts of the program on the stock of local public goods. This captures enduring improvements in the availability of functional agricultural drying floors, latrines, community centers and court "barries" (public buildings for dispute resolution). The measures of economic welfare suggest that one third of the initial gains dissipated over time. The remaining benefits reflect persistent increases in local market activity like the total number of petty traders and number of different goods for sale.[13] While we cannot speak directly to whether CDD outperforms other types of foreign aid or government provision, in our view these results are impressive, and particularly so given the challenges of working in a post-conflict environment.

Section VII: Expert Forecasts

Are these results about technocratic selection and the long-run effects of CDD in line with what we would have expected? To test this formally, we asked experts in both policy and academia to predict what we would find in our 2016 data collection with respect to both experiments. This exercise adds a few data points to broader efforts to systematically document prior beliefs and compare them to outcomes in the lab and the field (see Della Vigna and Pope 2018, Humphreys, Sanchez de la Sierra and van der Windt 2015, and Vivalt and Coville 2017). We asked experts to

[13] The lack of household survey data prevents us from directly measuring effects on wealth, consumption or income.

make predictions about community performance in the district government grants competition (for each of the six experimental cells in Figure 1), as well as long-run effects of CDD on the institutional and infrastructure measures collected in both 2009 and 2016.

We surveyed a variety of experts: policymakers in Sierra Leone with knowledge of the GoBifo project; policy experts working for multilateral aid agencies such as the World Bank, primarily based in Organization for Economic Co-operation and Development (OECD) countries; faculty in both economics and political science who have been directly involved in evaluating CDD projects or related areas of development (including ourselves, the co-authors of this article); and economics graduate students in Sierra Leone and OECD countries. Since we asked for several predictions in standard deviation units, which not all experts are familiar with, the survey first describes what these units mean, provides rules of thumb for what constitutes small versus large effects, and provides concrete examples of outcomes for each hypothesis (see instrument in Appendix B). We randomly varied whether the survey prompted the expert with the medium run results about CDD (from Casey et al. 2012). In total, we collected priors from 126 experts. Figure 4 summarizes their predictions along key dimensions (in panels A through C), where the circles denote individual expert predictions, the whisker plot portrays the distribution, and the realized effect size is presented with a solid horizontal line and accompanying 95% confidence interval.

Starting with long-run CDD impacts on institutions, it is apparent that policy experts in Sierra Leone greatly overestimated the potential for long-run institutional change, predicting average effects in the range of 0.25 standard deviation units. Policymakers in the OECD on average were roughly right on target. The predictions of graduate students closely track those of policymakers in their respective regions. While we cannot reject that economics and political science faculty were correct on average, they were more pessimistic: a substantial number of them (11 out of 23) predicted precisely zero long-run effects, which falls outside the 95% confidence interval of the observed point estimate. Pooled together, the experts predicted a long-run effect that significantly exceeds what was estimated in the medium run (0.095 predicted by experts, compared to 0.028 units in Casey et al. 2012), and remains statistically different from zero even when limited to the subgroup of experts who were primed with information on the short-run results. In our view, the substantial *ex ante* disagreement among seemingly well-informed experts about CDD's long-run institutional impacts makes this an interesting empirical exercise, and particularly so in light of the accumulation of shorter-run null results from several studies (see Wong 2012,

King and Samii 2014, White et al. 2017, and Casey 2018 for cross-country reviews).

This divergence between policymakers in Sierra Leone and academics lends some credence to concerns about optimism bias among policymakers and gripes (from policymakers) about hard-to-please academics, although note the substantial variation in priors among both types of expert. This potential disconnect does not appear to be as severe for policymakers based in the OECD countries, suggesting that the feedback loop between academic results and policy perceptions may be working fairly well for policymakers who are more proximate to rich country scholars, perhaps due to more frequent interactions at conferences and policy fora.

Expert predictions about the persistence of CDD-funded infrastructure investments were highly variable yet accurate on average. Pooled together, the experts predicted a long-run treatment effect of 0.218 standard deviation units (standard error 0.126), which is statistically indistinguishable from the estimated effect (of 0.204). Comparing priors for infrastructure to institutional outcomes in Figure 4, we see similar variation across expert types—policymakers in Sierra Leone were again relatively more optimistic about persistent gains and faculty more pessimistic—yet with even greater dispersion of predictions within each type of expert.

By contrast, all expert opinion diverged substantially from observed outcomes regarding entry into the infrastructure grants competition. We asked experts to make predictions about community entry into the competition for each cell in the technocratic selection experiment (in Figure 1). As a group, the experts predicted a baseline take up rate of 42 percent (for Arm 1), which reflects the sentiment of one expert who cautioned that "it is very likely that $2,500 is just too small an amount to get enough communities to bother with applying." In practice, we found a take up rate of 98%, which surprised all experts and far exceeded any prediction in the sample (Figure 4, Panel C). Experts also underestimated how much more effective technocratic selection would be when compared to CDD in helping communities avail of this new grants opportunity (using entry as a proxy for performance in the competition).

Specifically, comparing our two experimental interventions to each other, experts predicted that both CDD and technocratic selection would have a modest positive impact on community entry into the project challenge competition. On average, experts expected CDD and technocratic selection to boost community entry by 7 to 9 percentage points, for a 19 to 24 percent increase on the base rate predicted for chiefs in CDD control communities (see Appendix Table A1). Experts further predicted that training would enhance entry for technocrats by roughly 11 percentage

points, or an additional 22 percent. Note that we asked for priors about entry as opposed to proposal quality as we thought *ex ante* that it would be a meaningful margin of impact and an easy statistic for experts to engage with. *Ex post*, the unexpectedly high realized entry rates mean that quality would have been preferable. This disconnect reveals a methodological tension between locking in prior beliefs before results are known to avoid bias, and potential benefits of flexibility to adjust how priors are elicited after seeing the outcome data.

Taken together, these experiments offer a few data points on the question of when and how expert predictions may be useful in research: we see (i) disagreement across expert type for institutional change, (ii) wide dispersion for the durability of infrastructure, and (iii) systematic underestimation for both entry into the grants competition and the superiority of technocratic selection over CDD in this regard. While expert prior opinions may be useful for predicting some effects but not others, it remains unclear how to distinguish these cases *ex ante*. As more studies collect prior beliefs about the efficacy of policy interventions, a practice that is gaining some traction, we will be able to build a more thorough understanding of what types of impacts experts can predict, and which types of experts—those with country knowledge, practitioner experience or academic training—are most accurate.

Section VIII: Conclusion

Two randomized experiments suggest that encouraging communities to identify high skill residents and delegate technical aspects of local economic development projects to them holds promise as an effective and affordable strategy. In contrast, a long-running attempt to enhance participation in local governance and development projects yields little in the way of impacts on communities' ability to compete in the external grants competition that we study, as well as other dimensions of governance such as performance during the Ebola health crisis. These findings indicate that technocratic selection, accompanied by practical training in project management, may be a more viable and affordable strategy than attempts to affect institutional transformation in Sierra Leone.

The district government grants competition studied here provides a proof of concept for the idea that efforts to encourage delegation could unlock underutilized human capital, which could generalize to other areas of local governance. The "proof" lies in how clear the value of delegation seemed in this setting: grant writing is technical, requiring literacy and numeracy that

the chiefly elite generally do not possess, and failure to delegate decreases the odds of securing financial resources. That a majority of chiefs still failed to delegate in the status quo outlines the depth of the problem; and the high degree of responsiveness to objective information about skill and a nudge that encourages delegation outlines the potential. In rural Sierra Leone, other tasks that could be amenable to delegation to technocrats include securing funding and overseeing construction to build out the rest of the local infrastructure stock that is badly needed, managing recurrent budgeting and development planning efforts, and interpreting and applying government ordinances. Against a backdrop of statewide decentralization, the skills and talents of local managers will become increasingly important as greater authority transfers down from national politicians to local administrators.

In assessing external validity, note that impacts may have been quite different even if carried out in the same country just a decade earlier. When GoBifo launched, only 15% of adults had completed primary education and only 4% had completed secondary,[14] which would have greatly limited the scope for recruiting high skill residents in many villages. After the massive expansion of primary education in post-war Sierra Leone bolstered the human capital stock, there are many more skilled managers for communities to choose from, so long as local leaders are willing to consider younger, non-elite residents. As most low income countries in Africa and Asia have considerably better educated populations than Sierra Leone, similar forms of technocratic selection appear to be viable strategies in much of the world.

We are not able to directly test whether local institutions "matter" for development, as they proved quite resistant to a long-running reform effort in this setting. Yet in places where local democratization and other institutional reforms are not feasible, the question becomes moot from a policy perspective, and what we show here is that there exists a promising low cost alternative. Future research could usefully focus on other ways to improve institutions in light of the accumulated evidence that CDD programs are largely ineffective in this domain, as well as on other approaches to harness local skill even when institutional transformation is not practical.

[14] Casey et al. (2013) baseline household data from 2005: highest education level attained by all living household roster respondents (15 years and above).

References

Acemoglu, D. 2003. “Why not a political Coase theorem? Social conflict, commitment, and politics.” *Journal of Comparative Economics*, 31(4): 620-652

Acemoglu, D., S. Johnson and J. A. Robinson. 2001. “The Colonial Origins of Comparative Development: An Empirical Investigation.” *American Economic Review*, 91(5): 1369-1401.

Acemoglu, D., T. Reed, and J. A. Robinson. 2014. “Chiefs: Economic Development and Elite Control of Civil Society in Sierra Leone,” *Journal of Political Economy*, 122 (2): 319-368.

Alesina, A., and G. Tabellini. 2007. "Bureaucrats or Politicians? Part I: A Single Policy Task." *American Economic Review*, 97 (1): 169-179.

Alesina, A., and G. Tabellini. 2008. “Bureaucrats or Politicians? Part II: Multiple Policy Tasks.” *Journal of Public Economics* 92: 426-447.

Alva, C., C. Neilson, M. Bobba, M. Nieddu and G. Leon. 2017. “Teacher Wages, Student Achievement, and the Recruitment of Talent in Rural Peru,” working paper.

Anderson M. L., and J. Magruder. 2017. “Split-Sample Strategies for Avoiding False Discoveries.” Working Paper.

Ashraf, N., O. Bandiera, and S. Lee. 2016. “Do-gooders and Go-getters: Selection and Performance in Public Service Delivery.” Working paper.

Baldwin, K. 2016. *The paradox of traditional chiefs in democratic Africa*. New York City: Cambridge University Press.

Barro, R. 1991. “Economic Growth in a Cross Section of Countries,” *Quarterly Journal of Economics*, 106(2): 407-443.

Beath, A., F. Christia, and R. Enikolopov. 2013. “Do Elected Councils Improve Governance? Experimental Evidence on Local Institutions in Afghanistan.” MIT Political Science Department Research Paper No. 2013-24.

Besley, T., and M. Ghatak. 2005. "Competition and Incentives with Motivated Agents." *American Economic Review*, 95 (3): 616-636.

Bloom, N., B. Eifert, A. Mahajan, D. McKenzie and J. Roberts. 2013. “Does Management Matter? Evidence from India,” *Quarterly Journal of Economics*, 128 (1): 1-51.

Bloom, N., A. Mahajan, D. McKenzie and J. Roberts. 2018. “Do Management Interventions Last? Evidence from India,” NBER Working Paper No. 24249.

Bulte, E., P. Richards and M. Voors. 2018. *Institutions and Agrarian Development: A New*

Approach to West Africa, Palgrave Macmillan Publishers

Casey, K. 2018. "Radical Decentralization: Does Community Driven Development Work?" *Annual Review of Economics*, 10: 139-165.

Casey, K., R. Glennerster and E. Miguel. 2012. "Reshaping Institutions: Evidence on Aid Impacts Using a Pre-Analysis Plan," *Quarterly Journal of Economics*, 127(4):1755-1812.

Casey, K., R. Glennerster and E. Miguel. 2013. "Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan," https://hdl.handle.net/1902.1/21708, Harvard Dataverse.

Central Statistics Office. 1985. "The Analytical Report 1985 Census of Population and Housing Sierra Leone," Technical report.

Dal Bó, E., F. Finan and M. A. Rossi. 2013. "Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service," *Quarterly Journal of Economics*, 128(3): 1169-1218.

DellaVigna, S and D Pope. 2018. "What Motivates Effort? Evidence and Expert Forecasts," *The Review of Economic Studies*, 85(2): 1029–1069.

DellaVigna, S and D Pope. Forthcoming. "Predicting Experimental Results: Who Knows What?" *Journal of Political Economy*.

Deserranno, E. 2017. "Financial Incentives as Signals: Experimental Evidence from the Recruitment of Village Promoters in Uganda," Buffett Institute Global Poverty Research Lab Working Paper No. 17-105.

Dongier, P, Domelen JV, Ostrom E, Rizvi A, Wakeman W, et al. 2002. Chapter 9: Community-Driven Development. In *Volume 1 – Core Techniques and Cross-Cutting Issues*. World Bank, Washington D.C.

Fearon J, Humphreys M, Weinstein J. 2015. "How Does Development Assistance Affect Collective Action Capacity? Results from a Field Experiment in Post-Conflict Liberia." *American Political Science Review* 109(3):450-69.

Finan, F., B. Olken and R. Pande. 2017. "The Personnel Economics of the Developing State" in Handbook of Field Experiments, Volume II. North Holland: A. Banerjee, E. Duflo (eds).

Gennaioli, N., R. La Porta, F. Lopez-de-Silanes, A. Andrei Shleifer. 2013. "Human Capital and Regional Development," *Quarterly Journal of Economics*, 105-164.

Glaeser, E, R. La Porta. F. Lopez-de-Silanes, A. Shleifer. 2004. "Do Institutions Cause Growth?" NBER Working Paper 10568.

GoBifo Project. 2007. "Operations Manual: Version 2 June 2007." Project mimeograph.

He, G. and S. Wang. 2017. "Do College Graduates Serving as Village Officials Help Rural China?" *American Economic Journal: Applied Economics*, 9(4): 186-215.

Humphreys M., R. Sanchez de la Sierra and P. van der Windt. 2015. "Social Engineering in the Tropics: A Grassroots Democratization Experiment in the Congo." Working Paper.

Huntington, S. 1968. *Political Order in Changing Societies*. New Haven: Yale University.

King E, Samii C. 2014. "Fast-Track Institution Building in Conflict-Armed Countries? Insights from Recent Field Experiments," *World Development* 64:740-54.

Kling, J., J. Lieberman and L. Katz. 2007. "Experimental Analysis of Neighborhood Effects," *Econometrica*, 75(1): 83–119.

Kremer M, Miguel E. 2007. "The Illusion of Sustainability. *The Quarterly Journal of Economics*, 122(3):1007-65.

Mansuri G, Rao V. 2013. Localizing Development: Does Participation Work? *Policy Research Report*. Washington, D.C.: World Bank.

Martinez-Bravo, M, G. Padró i Miquel, N. Qian and Y. Yao. 2017 "The Rise and Fall of Local Elections in China: Theory and Empirical Evidence on the Autocrat's Trade-off," NBER working paper 24032.

McKenzie, D. and C. Woodruff. 2014. "What Are We Learning from Business Training and Entrepreneurship Evaluations around the Developing World?" *The World Bank Research Observer*, 29(1): 48–82.

Miguel, E., Gugerty MK. 2005. "Ethnic Diversity, Social Sanctions, and Public Goods in Kenya," *Journal of Public Economics* 89:2325-68.

North, D. C. 1990. *Institutions, Institutional Change and Economic Performance.* Cambridge: Cambridge University Press.

Rasul, I., and D. Rogger. 2016. "Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service," *Economic Journal*, 128: 413–446.

Reno, W. 1995. *Corruption and State Politics in Sierra Leone*. Cambridge and New York: Cambridge University Press.

Rodrik, D., A. Subramanian and F. Trebbi. 2004. "Institutions Rule: The Primacy of Institutions Over Geography and Integration in Economic Development," *Journal of Economic Growth*, 9(2): 131-165.

Richards, P. 1996. *Fighting for the Rainforest: War, Youth and Resources in Sierra Leone*.

Vivalt, E. and A. Coville. 2017 "How Do Policymakers Update?" Working Paper.

Voors, M., T. Turley, E. Bulte, A. Kontoleon, and J.A. List. 2018. "Chief for a Day: Elite Capture and Management Performance in a Field Experiment in Sierra Leone," *Management Science*

White, H. 1999. "Politicizing Development? The Role of Participation in the Activities of Aid Agencies." In *Foreign Aid: New Perspectives* ed. Kanhaya L, Gupta. Boston: Kluwwe Academic Publishers.

White, H, Menon R, Waddington H. 2017. "Community-driven development: does it build social cohesion or infrastructure? A mixed-method evidence synthesis report." 3ie Working Paper.

Wong S. 2012. "What Have Been the Impacts of World Bank Community-Driven Development Programs? CDD Impact Evaluation Review and Operational Research Implications." Washington D.C.: World Bank.

Woolcock, M. 2013. "Using Case Studies to Explore the External Validity of 'Complex' Development Interventions," *Evaluation*, 19(3): 229–248.

World Bank. 2017. "Education Profile for Sierra Leone," Africa Education Sector Unit for West and Central Africa Technical report.

Figure 1: Experimental Design

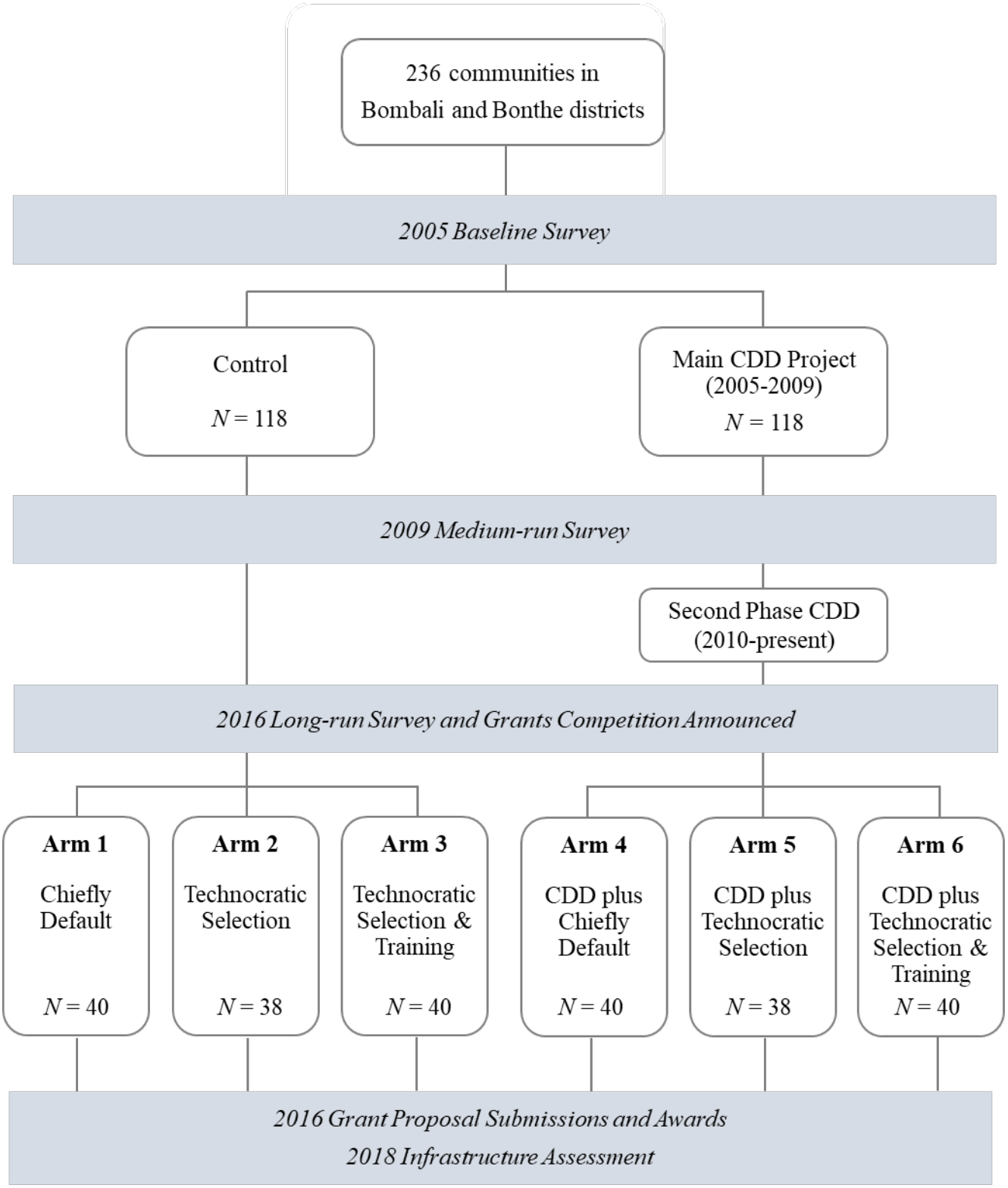
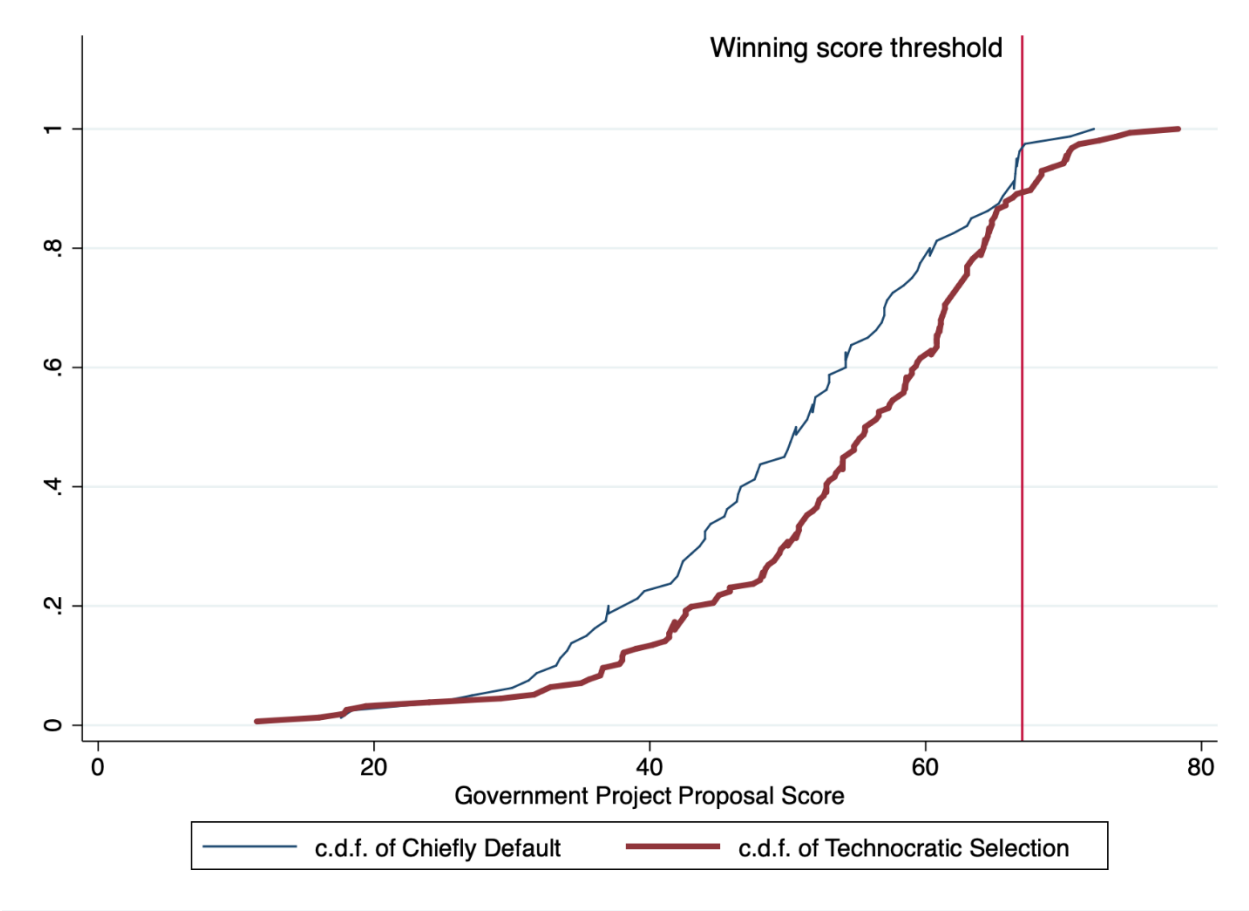
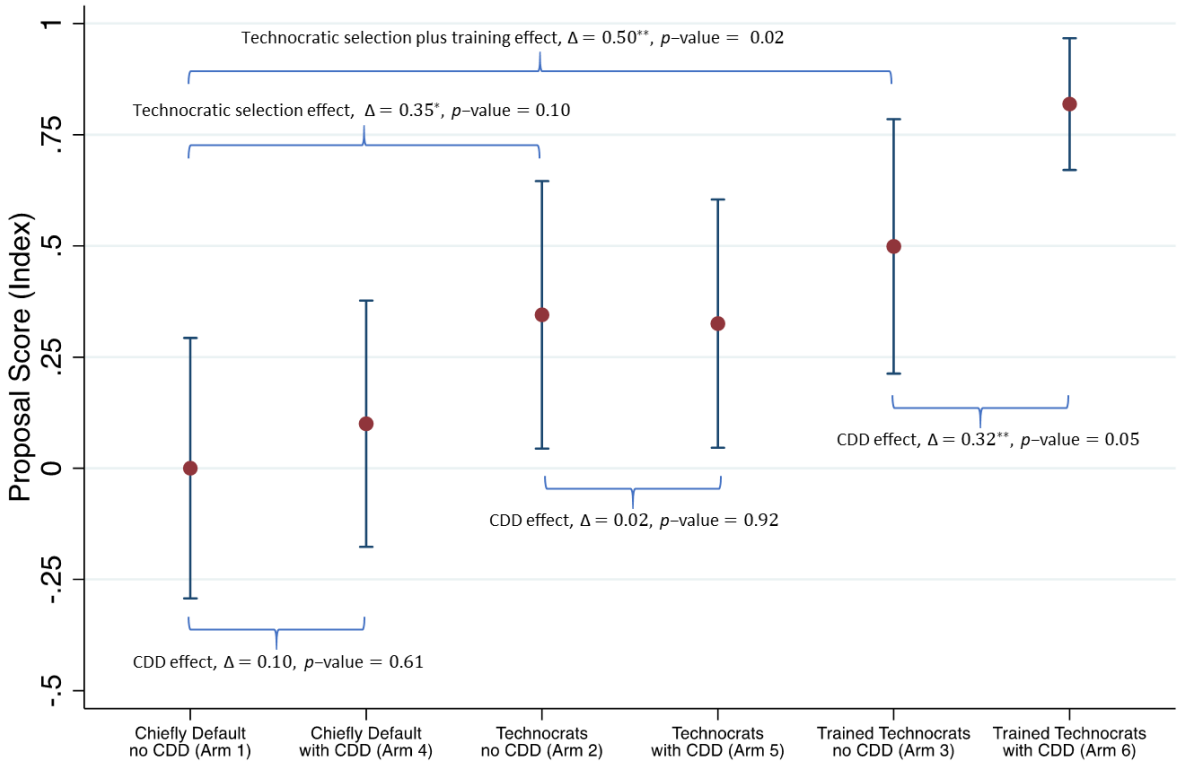


Figure 2: Distribution of Government Proposal Scores by Treatment Assignment



Notes: This figure presents the cumulative density of the scores the relevant district governments gave to proposals submitted by communities, separately for those assigned to the chiefly default condition (arms 1 and 4 in Figure 1) and to the technocratic selection treatment (arms 2, 3, 5 and 6). The vertical line demarcates the minimum score threshold that determines which communities won an implementation grant (standardized by minus 1 point for Bombali District to place both districts on a uniform scale). Scores imputed at experimental arm mean for the four non-submitting communities (N = 236). A Kolmogorov-Smirnov test rejects equivalence of the two distributions at p-value = 0.03.

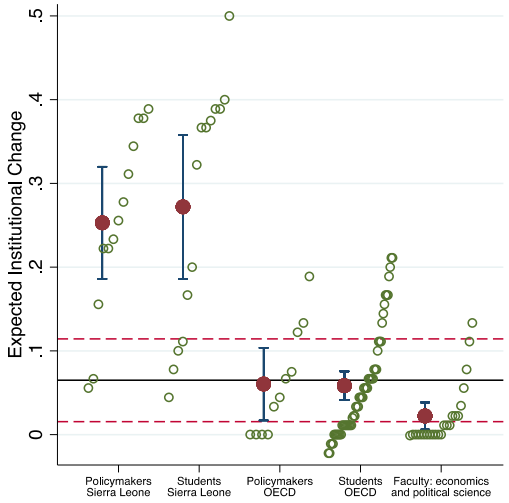
Figure 3: Proposal Performance across Manager Selection Arms



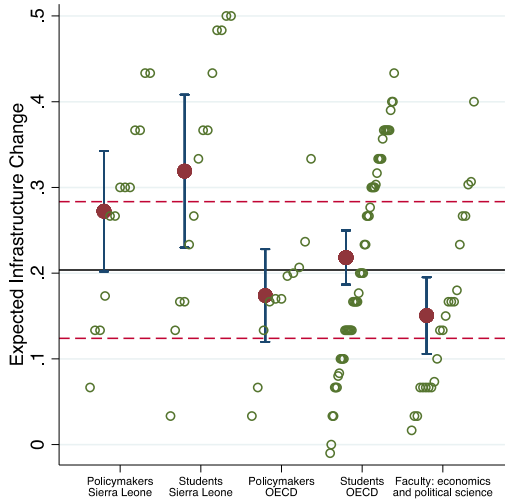
Notes: This figure presents the mean proposal score index and 95 percent confidence interval for the different types of managers in each of the six experimental arms indicated in Figure 1. Scores are standardized with respect to chiefs in the default condition without CDD exposure (Arm 1) and expressed in standard deviation units. The brackets compare two specific treatment arms to each other and report the difference in mean scores and associated p-value from a t-test of equality of means across arms. The positive and marginally significant difference between Arm 2 and Arm 1 above captures the "pure" effect of technocratic selection in the absence of CDD. The positive and highly significant difference between Arm 3 and Arm 1 captures the combined effect of selecting and training technocrats in the absence of CDD. The three brackets below the point estimates capture the effect of CDD across comparable arms in the technocratic selection experiment. The first two null results suggests that neither chiefs nor technocrats perform any better in CDD versus control communities. The rightmost bracket suggests that technocrats with CDD experience responded more strongly to the management training. Yet note that in the regression analogue (in Appendix Table A4), the F-test cannot reject that all three CDD estimates are jointly equal to zero, while the comparable F-test for the four technocratic selection and training arms rejects at above 99 percent confidence. Missing values for communities that did not submit a proposal are imputed at the relevant treatment arm mean.

Figure 4: Expert Predictions of Long-run CDD Effects and Grants Competition

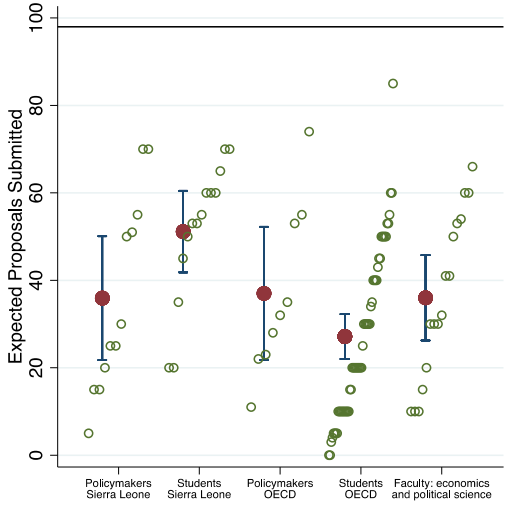
Panel A: Long-run Institutional Change



Panel B: Long-run Infrastructure Change



Panel C: Entry into the Grants Competition



Notes: This figure presents expert predictions collected during December 2016 and July 2017 before any data analysis. Panels A and B present expectations for CDD treatment effects measured in standard deviation units. The realized effect size is presented with solid black horizontal lines and the accompanying 95% confidence interval is demarcated by dashed horizontal lines. Panel C presents expectations about the percent of communities in the base case (no CDD, chiefly default condition, or Arm 1 of Figure 1) that would enter the grants competition. The realized point estimates are: i) 0.066 standard deviation unit (standard error 0.025) CDD treatment effect on institutions for Panel A; b) 0.204 standard deviation unit (standard error 0.040) CDD treatment effect for infrastructure in Panel B; and c) 98.3% percent of communities entered the grants competition for Panel C. For Panels A and B, expert predictions were closer to the realized value for the version of the survey that provided the short to medium run results for institutional change (p-value <0.01) but not statistically distinct for infrastructure (p-value=0.27).

Table 1: Treatment Effects on Performance in the Grants Competition

	Proposal Score (index)	Technical Score	Expert Score	Gov't Score	Won a Grant
	(1)	(2)	(3)	(4)	(5)
Panel A: Technocratic Selection versus CDD					
Technocratic Selection	0.397**	0.526***	0.377**	0.289	0.102**
	(0.164)	(0.193)	(0.169)	(0.177)	(0.049)
CDD	0.061	-0.015	0.063	0.136	0.049
	(0.181)	(0.206)	(0.192)	(0.190)	(0.047)
Technocratic Selection * CDD	0.094	0.017	0.218	0.047	-0.087
	(0.222)	(0.255)	(0.232)	(0.238)	(0.068)
Observations	236	236	236	236	236
F -statistic (on TS and TS*CDD)	8.00	8.65	9.01	3.44	2.17
p -value	<0.001	<0.001	<0.001	0.034	0.12
Panel B: Technocratic Selection and Managerial Training					
Technocratic Selection	0.315**	0.435***	0.298**	0.214	0.067
	(0.138)	(0.156)	(0.140)	(0.152)	(0.044)
Training	0.339**	0.280*	0.446***	0.292*	-0.013
	(0.133)	(0.157)	(0.130)	(0.155)	(0.049)
F -statistic (on TS and TR)	12.59	11.61	16.09	5.86	1.446
p -value	<0.001	<0.001	<0.001	0.003	0.238
Observations	236	236	236	236	236

*Notes: i) significance levels indicated by * p < 0.10, ** p <0.05, *** p < 0.01; ii) robust standard errors; iii) specifications in Panel A pool the technocratic selection and training arms together (see Appendix Table A4 for full interaction model) and include strata for geographic ward and two balancing variables (distance to road and community size) from the original randomization; iv) specifications in Panel B include the two balancing variables and strata for ward crossed with CDD assignment; v) outcomes in columns 2 to 4 are mean effects indices, expressed in standard deviation units, standardized with respect to the mean and standard deviation of control arm 1 (arms 1 and 4) in Figure 1 for Panel A (B) (see Kling, Liebman and Katz 2007); vi) missing scores for the 4 non-submitting communities are imputed at the respective treatment arm mean (see Appendix Tables A2 and A3 for imputation bounds); vii) outcome in column 1 is an equally weighted index of those in columns 2 to 4; viii) outcome in column 5 is a binary indicator; ix) Training term in Panel B captures the additional effect of training beyond that of technocratic selection; x) the F-statistic and associated p-value evaluate the hypothesis that the listed terms are jointly equal to zero; and xi) the sample for all specifications includes all communities in Figure 1.*

Table 2: Variation in Characteristics of Managers and Community Nominees

Panel A: Chiefs versus Top-scoring Technocrats (in all communities)

	Chiefs	Technocrats	*p*-value on difference	*N*
Average age	58.04	37.77	<0.01	455
Proportion male	0.98	0.95	0.09	466
Proportion with any formal education	0.35	0.98	<0.01	468
Proportion born in this community	0.95	0.76	<0.01	468
Proportions in occupation groups:				
farmer	0.88	0.32	<0.01	468
teacher	0.01	0.44	<0.01	468
business (e.g. petty trading)	0.04	0.05	0.66	468
Score on managerial capital test	31.47	74.77	<0.01	468

Panel B: Technocratic Nominees in CDD Treatment versus Control Communities

	CDD Controls (arms 1-3)	CDD Treatment (arms 4-6)	*p*-value on difference	*N*
Average age	38.23	40.32	0.02	1,148
Proportion male	0.76	0.76	0.77	1,162
Proportion with any formal education	0.68	0.70	0.50	1,168
Proportion born in this community	0.76	0.80	0.10	1,168
Proportions in occupation groups:				
farmer	0.62	0.56	0.08	1,168
teacher	0.15	0.17	0.56	1,168
business (e.g. petty trading)	0.06	0.07	0.64	1,168
Score on managerial capital test	43.96	45.38	0.49	1,155

Notes: Panel A compares characteristics of the chief to the single highest scoring technocratic nominee in each community; Panel B compares the average characteristics of all five technocratic nominees in CDD treated versus control communities.

Table 3: Variation in Chief's Role in Project Management

Panel A: Technocratic Selection Effect	Chiefly Default (arms 1, 4)	Technocratic Selection (arms 2, 3, 5, 6)	*p*- value on difference
Proportion where chiefly authorities chose the project	0.80	0.45	<0.01
Proportion where chiefly authorities wrote the description	0.40	0.26	0.03
Proportion where chiefly authorities did the budget	0.37	0.22	0.02
Proportion where chiefly authorities set the timeline	0.38	0.26	0.07
Observations	221		
Panel B: CDD Effect in Full Sample	CDD Controls (arms 1-3)	CDD Treatment (arms 4-6)	*p* -value on difference
Proportion where chiefly authorities chose the project	0.64	0.51	0.08
Proportion where chiefly authorities wrote the description	0.32	0.28	0.49
Proportion where chiefly authorities did the budget	0.28	0.26	0.79
Proportion where chiefly authorities set the timeline	0.32	0.28	0.49
Observations	221		
Panel C: CDD Effect in Technocratic Selection Arms	CDD Controls (arms 2, 3)	CDD Treatment (arms 5, 6)	*p* -value on difference
Proportion where chiefly authorities chose the project	0.55	0.37	0.04
Proportion where chiefly authorities wrote the description	0.27	0.25	0.78
Proportion where chiefly authorities did the budget	0.23	0.22	0.91
Proportion where chiefly authorities set the timeline	0.28	0.25	0.65
Observations	148		

Notes: i) outcomes capture the propotion of management decisions that were made by the village headman or other chiefly authorities in the community; ii) Panel A compares communities assigned to technocratic selection (with or without training) to the default of chiefly control; iii) Panel B compares communities assigned to CDD treatment versus control; and iv) Panel C compares CDD treated versus control communities in the technocratic selection (with or without training) arms, to look at compliance with the assignment to delegate to technocrats.

Table 4: Long-run CDD Treatment Effects

	Treatment effect 2016 (1)	Naïve *p* -value (2)	FDR *q* -value (3)	Treatment effect 2009 (4)	Change over time (1) - (4)
Panel A: Institutions Family					
All outcomes (61 unique outcomes)	0.066***	<0.01	0.005	0.028	0.037
	(0.025)			(0.020)	(0.028)
Collective action	0.098	0.049	0.235	0.012	0.086
	(0.050)			(0.037)	(0.061)
Inclusion	0.033	0.350	0.539	0.002	0.031
	(0.036)			(0.032)	(0.044)
Local authority	-0.035	0.604	0.632	0.056	-0.088
	(0.068)			(0.037)	(0.070)
Trust	0.107	0.063	0.235	0.042	0.064
	(0.057)			(0.046)	(0.081)
Groups and networks	0.149	0.037	0.235	0.028	0.121
	(0.071)			(0.037)	(0.074)
Access to information	-0.036	0.590	0.632	0.038	-0.075
	(0.067)			(0.037)	(0.072)
Participation in local governance	0.079	0.190	0.348	0.090**	-0.011
	(0.060)			(0.045)	(0.065)
Crime and conflict	-0.002	0.971	0.759	0.01	-0.012
	(0.063)			(0.043)	(0.074)
Political and social attitudes	0.154	0.215	0.348	0.041	0.113
	(0.124)			(0.043)	(0.126)
Panel B: Infrastructure Family					
All outcomes (30 unique outcomes)	0.204***	<0.01	0.001	0.298***	-0.094***
	(0.040)			(0.031)	(0.036)
Project implementation	0.253***	<0.01	0.001	0.703***	-0.450***
	(0.068)			(0.055)	(0.081)
Local public goods	0.228***	<0.01	0.001	0.204***	0.024
	(0.046)			(0.039)	(0.041)
Economic welfare	0.240***	<0.01	0.001	0.376***	-0.136**
	(0.056)			(0.047)	(0.062)
Observations	236				

*Notes: i) significance levels indicated by * p < 0.10, ** p < 0.05, *** p < 0.01 based on FDR-adjusted q-values in column 1 and naive per comparison values in columns 4 and 5; ii) specifications include strata for geographic ward and two balancing variables (distance to road and community size) from the original randomization; iii) robust standard errors; iv) all estimates are for hypothesis-level equally weighted mean effects indices, expressed in standard deviation units (see Kling, Liebman and Katz 2007); v) the dependent variable in column 5 is the difference in 2009 and 2006 indices, where the set of component measures varies across survey round (see Appendix G for exact panel specification); and vi) 2009 data sourced from Casey et al (2012).*

ONLINE APPENDIX: MATERIAL NOT INTENDED FOR PUBLICATION

Contents:

Appendix A: Technocratic Selection Implementation Script

Enumerator A SCRIPT: Project Challenge and Manager Selection

STEP 1: Explain project challenge

READ TO GROUP: The Local Councils in Bombali and Bonthe are running a new exciting project challenge competition in your area. They are asking communities to submit proposals for small scale infrastructure (like construction of a latrine or drying floor, or repairs to a local school building). The Councillors will evaluate the proposals and pick the 20 best proposals as the winners. These 20 winning communities will receive **14 Million Leones** to use for implementing their projects. This is a lot of money! Your community is eligible to participate and I would like to encourage you to apply.

[HOLD UP THE PROPOSAL FORM FOR ALL TO SEE] This is the proposal form you will need to fill out to enter the project competition. I want this community to do well in this competition so will explain the things you need to put into a proposal and ask you to think about people in this community who would be good at putting these things together.

First, a strong project proposal needs a clear **description of the project**. This section tells the Council what the project will be, why the project solves an important problem or addresses an urgent need, and who will benefit from the project. To develop this description, you need a project leader who is good at identifying problems, coming up with solutions, making a persuasive argument ("sabi tok"), and who can read and write well.

Second, a winning project proposal needs to have a clear and reasonable **budget**. The budget lists all the items you will need to construct the project, how much they will cost, and where you will get them. It needs to show that your project will deliver value for money. You need a project leader who is familiar with these kinds of construction projects, knows where to get things, and how to get them at a good price, and someone who is good with numbers.

Third, a strong project proposal sets out a clear plan of work and **timeline**. This part of the proposal tells the Council who will do what and when. It should show that you know how to get things done: you can mobilize the workers you need, or know how to find a good contractor to work for you. You need a project leader who can set deadlines for each part of the project and get things done on time.

Before we leave today we will give you this project application form that you can use to submit the proposal. We will also tell you the date before which you need to submit the proposals. The proposals should be submitted in person to the District Council office in Makeni/Mattru Jong.

The winners announcement will be done in January 2017. You will receive an invitation to participate in the awards ceremony. We hope you will apply!

STEP 2: Ask for nominations/volunteers

READ TO GROUP: Now I would like all of you to think about people in this village who are good at doing the things needed to develop a strong project proposal. I will step away from the group and let you think and talk about who would be good for this important job. We all know that the village headman has lots of experience running projects in this community. I would like you to also give me the names of 5 other people (in addition to the headman) that have these skills: they can **read and write**, they can come up with a **persuasive plan**, they know how to put a **budget** together, they are good at setting a **timeline**, **meeting deadlines** and **getting things done**. I will step away now so please call to me to come back when you have come up with the 5 people plus the headman.

STEP 3: Observe the proceedings

Step away outside the circle of the focus group and observe what happens.
Enumerator A: Fill out TALLY SHEET A below.
Enumerator B: Fill out TALLY SHEET B below.

STEP 4: Collect names of nominees / volunteers

Enumerator A: *When the community has finished its deliberation, rejoin the focus group and ask them to give you the names of the people they recommend.*

Name of Headman:_____
Name of 1st nominee:_____
Name of 2nd nominee:_____
Name of 3rd nominee:_____
Name of 4th nominee:_____
Name of 5th nominee:_____

NOTE: *if fewer than 5 nominees (in addition to the headman) were identified, only give the tests to the individual(s) selected by the focus group. If more than 5 nominees (in addition to the headman) were identified, ask the participants to rank the individuals and only work with the top 5 (plus the headman).*

READ TO GROUP: Thank you for these nominations. I would like to now ask each of these 5 nominated people to complete a short survey with me in private. The survey includes a test to measure the skills we talked about that are important for leading the project proposal: writing, making a project plan, doing a budget, working with numbers. The test will be done in private and the results will not be made public. Once all the tests are done, we will come back together as a group and I will unlock the project leader lottery. This lottery will randomly pick who will be the project proposal leader: it will tell us whether the leader for this project challenge competition will be A) the person with the highest score on the management test; or B) the village headman. I myself do not know which person the lottery will pick, and I cannot unlock the lottery until everyone completes the test. So let us please take a break and come back together at [TIME] to unlock the lottery and see who will lead

the project challenge competition for this village!

STEP 5: COMPLETE THE MANAGEMENT TESTS

Complete the management tests with all 6 people above. Score the tests on site IN PRIVATE. When finished, see which person of the 5 NON-HEADMAN nominees had the highest score on the test. Make sure you know this person's name so you can announce it to the group if the lottery picks the HIGHEST SCORER to be the project leader. Do NOT share any information on how people scored on the management test.

STEP 6: RECONVENE THE FOCUS GROUP TO UNLOCK THE LOTTERY

READ TO GROUP: Thank you for coming back together. We can now unlock the project leader lottery! Remember, it will randomly pick whether A) the person with the highest score on the management test or B) the village headman will be the leader for the project challenge competition.

[UNLOCK THE LOTTERY: HOLD THE SCREEN UP SO THAT EVERYONE CAN SEE THE LOTTERY RUNNING. ANNOUNCE THE LOTTERY RESULT TO THE GROUP]

STEP 7: NEXT STEPS VARY BY LOTTERY RESULT

➔ **IF THE LOTTERY SAYS "HEADMAN LEADER":**

Explain that the lottery has randomly chosen the HEADMAN to be in charge of the project proposal for the challenge competition. Show the group the project application form and say that you are writing the HEADMAN down as the project proposal leader. Write his name on the application in front of the group. Walk over to the HEADMAN and give him the project application form. Explain that the proposal should be submitted in person by himself. Also give him the transportation voucher and explain that this can be redeemed when the proposal is submitted. Tell him that you hope he will put together a proposal for this village and that he will submit it to the Local Council.

Announce that the proposal needs to be submitted to [LOCAL COUNCIL ADDRESS] before the deadline [DATE]. Encourage them to apply.

Thank everyone for their time and wish them good luck with the project challenge competition!

END MEETING HERE AND GO TO VILLAGE INSPECTION SES SURVEY SECTION N

➔ **IF THE LOTTERY SAYS "HIGHEST SCORER":**

Explain that the lottery has randomly chosen the person with the highest management test score to be in charge of the project proposal for the challenge competition. Remind the group that you have used some tests to measure the skills needed for a strong proposal—reading

and writing, budget and costing, previous project experience—and that the tests have identified [NAME OF HIGHEST SCORER] as the person with the strongest skills for this particular opportunity. Show the group the project application form and say that you are writing [NAME OF HIGHEST SCORER] down as the project proposal leader. Write his name on the application in front of the group. Walk over to [NAME OF HIGHEST SCORER] and give him/her the project application form. Explain that the proposal should be submitted in person by the [NAME OF HIGHEST SCORER]. Also give him/her the transportation voucher and explain that this can be redeemed when the proposal is submitted. Tell him/her that you hope he/she will put together a proposal for this village and submit it to the Local Council.

Announce that the proposal needs to be submitted to [LOCAL COUNCIL ADDRESS] before the deadline [DATE]. Encourage them to apply.

Thank everyone for their time and wish them good luck with the project challenge competition!

END MEETING HERE AND GO TO VILLAGE INSPECTION SES SURVEY SECTION N

➔ **IF THE LOTTERY SAYS "HIGHEST SCORER + TRAINING":**

Explain that the lottery has randomly chosen the person with the highest management test score to be in charge of the project proposal for the challenge competition. Remind the group that you have used some tests to measure the skills needed for a strong proposal—reading and writing, budget and costing, previous project experience—and that the tests have identified [NAME OF HIGHEST SCORER] as the person with the strongest skills for this particular opportunity. Show the group the project application form and say that you are writing [NAME OF HIGHEST SCORER] down as the project proposal leader. Write his name on the application in front of the group. Walk over to [NAME OF HIGHEST SCORER] and give him/her the project application form. Explain that the proposal should be submitted in person by the [NAME OF HIGHEST SCORER]. Also give him/her the transportation voucher and explain that this can be redeemed when the proposal is submitted. Tell him/her that you hope he/she will put together a proposal for this village and submit it to the Local Council.

Announce that the proposal needs to be submitted to [LOCAL COUNCIL ADDRESS] before the deadline [DATE].

READ TO GROUP: And, this village is very fortunate as you have qualified for a special one day training session that the Local Councils are offering in your area to teach you how to develop a successful project proposal. The session will cover the critical steps we discussed earlier: how to write a project description, how to draft a budget and how to set and meet deadlines, plus many other useful skills. I want to be sure that this village benefits from this training so will also cover the transport costs of [NAME OF HIGHEST SCORER] to participate in this important training.

Give [NAME OF HIGHEST SCORER] the TRAINING voucher that can be redeemed for full

transport costs plus food and drinks at the training.

Announce that the training session will be held at [LOCATION] on this day [DATE] at this time [TIME]. Encourage them to [NAME OF HIGHEST SCORER] to attend the training!

Thank everyone for their time and wish them good luck with the project challenge competition!

END MEETING HERE AND GO TO VILLAGE INSPECTION SES SURVEY SECTION N

Appendix B. Expert Prior Elicitation Details

Before collecting and analyzing the data, we first established what experts in the field *thought* we would find. To do so, we fielded a survey among different types of experts and asked them to make predictions in two main areas: i) long run impacts of CDD on measures of institutions and local public goods; and ii) community performance in the infrastructure grants competition for all six experimental cells in Figure 1.

Experts came from several groups: i) policymakers working for multilateral aid agencies (including the World Bank, the Department for International Development, the United Nations Development Programme and the International Rescue Committee) located mostly in OECD countries; ii) policymakers in Sierra Leone with knowledge of the GoBifo project; iii) economics graduate students in the United States (at University of California, Berkeley) and the Netherlands (at Wageningen University); iv) economics undergraduate students in Sierra Leone (at Fourah Bay College); and v) faculty in economics and political science directly involved in evaluating CDD projects (including the co-authors of this study) and other development economics researchers. This yielded 126 completed surveys in total, composed of 25 surveys from policymakers (12 in the OECD and 13 in Sierra Leone), 78 from students (17 undergraduate and 61 graduate students), and 23 from faculty. Survey response rates were quite high for all groups (e.g. 84% for faculty and 99% for graduate students) save the OECD policymakers (39% completion).

For estimates about long run CDD impacts, we used the same twelve hypotheses and comparable empirical measures that we focused on in Casey et al. 2012. For each hypothesis, we asked experts to predict the point estimates we would find in the long-run, in standard deviation units, and also indicate their level of certainty for each prediction (following DellaVigna and Pope 2018, forthcoming). As in our earlier work, we then group these hypotheses and predictions into two main families, infrastructure and institutions. There were two versions of the survey: the first provided detailed information on our medium run results and the second asked the expert to make predictions without any information provided (see instrument on page A9). We randomized which version was given to each expert, with a few exceptions (e.g. a small subset completed both versions). Expert predictions about the infrastructure grants competition focus on entry as a proxy for overall performance. Table A1 below presents results for each experimental cell in Figure 1.

Table A1: Predicted Entry into Grants Competition by Experimental Arms

	2005 Assignment:	
2016 Assignment:	CDD Control	CDD Treatment
Chiefly Default	*Arm 1*	*Arm 4*
	35.5%	42.2%
	(23.0)	(21.1)
Technocratic Selection	*Arm 2*	*Arm 5*
	44.0%	53.9%
	(22.3)	(20.7)
Trained Technocrats	*Arm 3*	*Arm 6*
	53.6%	65.5%
	(23.5)	(20.9)
Realized entry, all communities:	98.3%	

Notes: This table presents mean expert predictions about the percent of communities that would enter the project challenge competition in each of the six distinct treatment arms in Figure 1. We pool predictions across all 118 experts, who were surveyed between December 2016 and July 2017, before data analysis.

Expert Survey Instrument[1]:
Measuring the Long-Run Effects of Community Driven Development in Sierra Leone

Researchers: Katherine Casey, Rachel Glennerster, Edward Miguel, and Maarten Voors

Date: [Month, Year]

Overview: In 2012, we published the results of an impact evaluation of a community driven development (CDD) project in rural Sierra Leone, called GoBifo. That paper focused on the medium-run effects of CDD on local economic and institutional outcomes. We now plan to implement a new research project to measure the long-run effects of that project. Before we do so, we would value your input regarding what you expect these impacts to be, and have therefore prepared this brief (roughly 10 minute) survey.

Your participation is completely voluntary and you are free to leave the survey blank if you do not wish to participate. We will maintain your confidentiality by not recording any personally identifying information about you. We foresee little benefit or risk from participation, and cannot and do not guarantee or promise that you will receive any benefits from this study. If you have any questions about this research, please contact Katherine Casey at +1 (650) 725-2167. If you have any complaints, please contact the Stanford Human Subjects Institutional Review Board (IRB) at +1 (866) 680-2906.

1. **What is your job/position title?** _____

1. **What is your major?** _____

1 . What best describes your professional position and experience? (CIRCLE ONE)

a. Researcher who has worked on CDD evaluations
b. Researcher who has not worked directly on CDD
c. Development practitioner who has worked on implementing CDD projects
d. Development practitioner who has not worked directly on CDD

2. **Have you heard about the project challenge competition currently running in Bombali and Bonthe?** (CIRCLE ONE)
YES / NO

2. **In what year of your program are you?** _____

2. **Do you have any direct professional experience in Sierra Leone?** (CIRCLE ONE) YES / NO

3. **On a scale of 1 to 10, how familiar are you with <u>our 2012 study</u> of a CDD project in Sierra Leone entitled**
"Reshaping Institutions: Evidence on Aid Impacts Using a Pre-analysis Plan"
(with 1 representing having never heard of it to 10 being very familiar with the results)? (CIRCLE ONE)

1 ------- 2 ------- 3 ------- 4 ------- 5 ------- 6 ------- 7 ------- 8 ------- 9 ------- 10
Never heard of it *Very familiar with results*

4. **On a scale of 1 to 10, how familiar are you with <u>other CDD impact evaluations</u> in low income countries**

[1] Note that a few different versions of the survey were implemented. This version includes priming with the medium run results, the other version omitted these primes but was otherwise the same. Color coded text mark small differences in questions across pools of expert, where (i) **Black** is universal except questions 1 and 2 was given to Academic Experts, Policy Experts, and the PIs; (ii) Blue was given to students in Sierra Leone and Berkeley; and (iii) Red was given to Sierra Leone Policy Makers and Wageningen students.

(with 1 representing having never heard about other CDD studies to 10 being very familiar with the results of several studies)? (CIRCLE ONE)

1 ------- 2 ------- 3 ------- 4 ------- 5 ------- 6 ------- 7 ------- 8 ------- 9 ------- 10

Never heard of any *Very familiar with several*

5. **Do you think that the World Bank <u>should continue to support</u> community driven development (CDD) programs to the extent that it currently does?** (CIRCLE ONE)
    a. The World Bank should spend more on CDD than current amount
    b. The World Bank should maintain current levels of spending
    c. The World Bank should spend less on CDD than current amount
    d. Indifferent

Standard Deviation Unit Effect

In what follows, we will ask you to predict how large the long-run treatment effects of the Sierra Leone CDD project will be. As we measure effects across groups of outcomes, standard practice is to refer to treatment effect sizes in <u>standard deviation units (sdu's)</u>. This makes the effect sizes comparable across outcome measures. For your reference, the following table provides a rule of thumb interpretation of the real-world magnitude of standard deviation unit treatment effects of various sizes (in absolute value):

Treatment effect size in standard deviation units (sdu's), in absolute value	**Interpretation**
0.00	No impact
0.05	Very small effect
0.10	Small effect
0.20	Moderately small effect
0.30	Moderate effect
0.40	Moderately large effect
> 0.50	Large effect

PART I. MEDIUM-RUN RESULTS AND LONG-RUN FORECASTS

The CDD Project "GoBifo" (which means "move forward" in the dominant local language) in Sierra Leone was implemented from 2005 to 2009. This project provided block grants of US$5,000 (approximately US$100 per household) to communities in rural Sierra Leone. The grants could be used for the construction of local public goods, trade skills training, and small business start-up capital. GoBifo facilitators spent an average of 6 months in each of these villages promoting democratic decision-making, the participation of socially marginalized groups (such as women and youth), and transparent local budgeting practices. In addition, 60 of these villages received a follow up grant of $1,300 in 2010 for youth empowerment programs.

The project was implemented as a randomized control trial, where 118 villages participated in the GoBifo intervention and 118 served as controls that did not receive any project assistance. The original follow-up survey of medium-run treatment effects was fielded in 2009 and evaluated impacts on 12 hypotheses which we grouped into two broad sets of indicators: a family of "hardware" effects on local public goods and economic outcomes, and a family of "software" effects including institutional and social capital measures. We are now going back to the field to measure long-run effects, a full 7 years after the program ended, and would like to know your views on what you expect the long-run effects of GoBifo are likely to be.

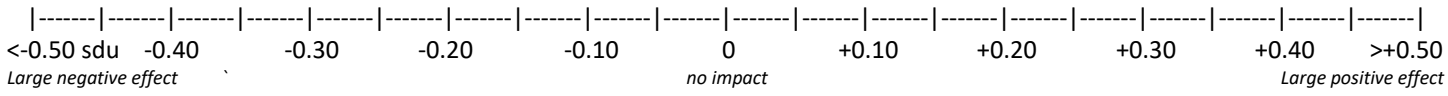
Since there are several individual outcome measures included under each of the 12 hypotheses, we measure the average effect across all of them after normalizing measures in standard deviation units (sdu's). Below we list all 12 hypotheses tested in the study and include examples of indicators used in the survey. We also provide you with detailed results from our 2012 study of the medium-run effects of the GoBifo project.

For each of 12 hypotheses below, please mark the scale with an **X** for the size of the **long-run treatment effect** of the GoBifo project that you expect we will find when we return to the field in Sierra Leone to collect data this November. We would now like to provide you more detailed results from our 2012 study of the medium-run effects of the GoBifo project, and ask you to again predict what you think the long run effects of GoBifo will be for the following hypotheses.

Hardware family of outcomes

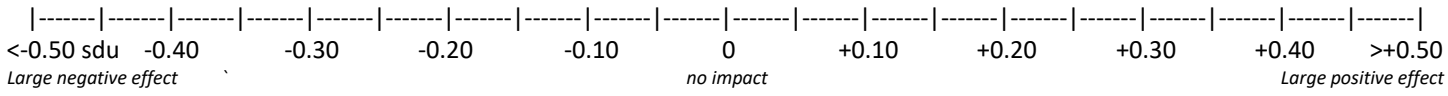
Hypothesis 1: GoBifo Project Implementation.
Examples of indicators include the presence of a village development committee and formal bank account for village project expenses.

Our study found medium-run effects for this hypothesis equal to **+0.70 sdu's,** which is statistically different from zero with a very high degree of confidence. What do you think the long run treatment effect will be?



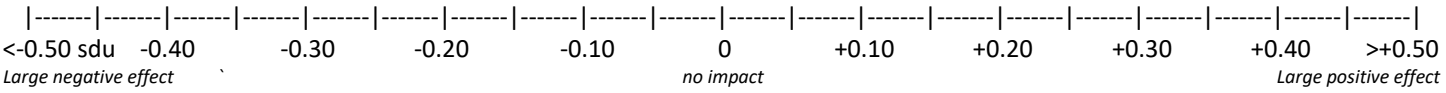
Hypothesis 2: Participation in GoBifo improves the quality of local public services infrastructure.
Examples include the presence and construction quality of latrines and drying floors.

Our study found medium-run effects equal to **+0.20 sdu's,** which is statistically different from zero with a very high degree of confidence. What do you think the long run treatment effect will be?



Hypothesis 3: Participation in GoBifo improves general economic welfare.
Indicators include the number of petty traders and goods on sale in the community.

Our study found medium-run effects for this hypothesis equal to **+0.38 sdu's,** which is statistically different from zero with a very high degree of confidence. What do you think the long run treatment effect will be?



Software family of outcomes

Hypothesis 4: Participation in GoBifo increases collective action and contributions to local public goods.
Indicators include presence of communal farms and community-supported teachers.

Our study found medium-run effects for this hypothesis equal to **+0.01 sdu's,** which is not statistically different than zero at traditional confidence levels. What do you think the long run treatment effect will be?

|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
<-0.50 sdu -0.40 -0.30 -0.20 -0.10 0 +0.10 +0.20 +0.30 +0.40 >+0.50
Large negative effect ` *no impact* *Large positive effect*

Hypothesis 5: GoBifo increases inclusion and participation in community planning and implementation, especially for poor and vulnerable groups; GoBifo norms spill over into other types of community decisions, making them more inclusive, transparent, and accountable.
Indicators include taking minutes at community meetings and reporting having fewer problems with financial misconduct.

Our study found medium-run effects equal to **0.00 sdu's,** which is not statistically different than zero at traditional confidence levels. What do you think the long run treatment effect will be?

|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
<-0.50 sdu -0.40 -0.30 -0.20 -0.10 0 +0.10 +0.20 +0.30 +0.40 >+0.50
Large negative effect ` *no impact* *Large positive effect*

Hypothesis 6: GoBifo changes local systems of authority, including the roles and public perception of traditional leaders versus elected local government.
Indicators include the community choosing a village headman younger than 35 years old.

Our study found medium-run effects equal to **+0.06 sdu's,** which is not statistically different than zero at traditional confidence levels. What do you think the long run treatment effect will be?

|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
<-0.50 sdu -0.40 -0.30 -0.20 -0.10 0 +0.10 +0.20 +0.30 +0.40 >+0.50
Large negative effect ` *no impact* *Large positive effect*

Hypothesis 7: Participation in GoBifo increases trust.
Indicators include the presence of cooperative trading groups that span multiple households.

Our study found medium-run effects for this hypothesis equal to **+0.04 sdu's,** which is not statistically different than zero at traditional confidence levels. What do you think the long run treatment effect will be?

|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
<-0.50 sdu -0.40 -0.30 -0.20 -0.10 0 +0.10 +0.20 +0.30 +0.40 >+0.50
Large negative effect ` *no impact* *Large positive effect*

Hypothesis 8: Participation in GoBifo builds and strengthens community groups and networks.
Indicators include presence of fishing groups / cooperatives in the community.

Our study found medium-run effects for this hypothesis equal to **+0.03 sdu's,** which is not statistically different than zero at traditional confidence levels. What do you think the long run treatment effect will be?

|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
<-0.50 sdu -0.40 -0.30 -0.20 -0.10 0 +0.10 +0.20 +0.30 +0.40 >+0.50
Large negative effect ` *no impact* *Large positive effect*

Hypothesis 9: Participation in GoBifo increases access to information about local governance.
Indicators include visits by local government officials and display of government policies or posters in the community.

Our study found medium-run effects equal to **+0.04 sdu's,** which is not statistically different than zero at traditional confidence levels. What do you think the long run treatment effect will be?

|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
<-0.50 sdu -0.40 -0.30 -0.20 -0.10 0 +0.10 +0.20 +0.30 +0.40 >+0.50
Large negative effect ` *no impact* *Large positive effect*

Hypothesis 10: GoBifo increases public participation in local governance.
Indicators include the involvement of local government officials in planning or overseeing community development projects.

Our study found medium-run effects equal to **+0.09 sdu's,** which is statistically different than zero with a moderate degree of confidence. What do you think the long run treatment effect will be?

|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
<-0.50 sdu -0.40 -0.30 -0.20 -0.10 0 +0.10 +0.20 +0.30 +0.40 >+0.50
Large negative effect ` *no impact* *Large positive effect*

Hypothesis 11: By increasing trust, GoBifo reduces crime and conflict in the community.
Indicators include reports of theft of household items or livestock.

Our study found medium-run effects for this hypothesis equal to **+0.01 sdu's,** which is not statistically different than zero at traditional confidence levels. What do you think the long run treatment effect will be?

|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
<-0.50 sdu -0.40 -0.30 -0.20 -0.10 0 +0.10 +0.20 +0.30 +0.40 >+0.50
Large negative effect ` *no impact* *Large positive effect*

Hypothesis 12: GoBifo changes political and social attitudes, making individuals more liberal towards women, more accepting of other ethnic groups and "strangers" and less tolerant of corruption and violence.
Indicators include community choosing a woman to be the village chief.

Our study found medium-run effects for this hypothesis equal to **+0.04 sdu's,** which is not statistically different than zero at traditional confidence levels. What do you think the long run treatment effect will be?

|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
<-0.50 sdu -0.40 -0.30 -0.20 -0.10 0 +0.10 +0.20 +0.30 +0.40 >+0.50
Large negative effect ` *no impact* *Large positive effect*

Overall expectations
You made 12 additional forecasts above about the long-run effects of GoBifo. How many of these additional forecasts do you think will fall within 10% of the true effect size (in standard deviation unit terms) that we find in the data we will begin to collect in November? _____ (out of 12)

PART II: FORECASTING THE RESULTS OF A NEW EXPERIMENT

The GoBifo project now plans to offer a lower-cost alternative to CDD that gives communities simple tools to help them identify people within the village with high managerial capital and provide these individuals with basic project management training. We will evaluate the effectiveness of this approach via a randomized experiment that we launched this Fall. We are interested in your forecasts about how effective these interventions will be.

As background, all 236 communities in our Sierra Leone study sample will be eligible **to submit a proposal for a grant for small-scale public infrastructure** as part of a new project challenge competition that the local governments are running. Local governments will award US$2,500 implementation grants to the twenty highest quality proposals. Two new types of support will be offered to increase community take up of this opportunity: i) simple tools to help communities identify individuals who have the skills (e.g., literacy and numeracy skills) necessary to lead a successful proposal development and project implementation, and ii) basic training in project management skills.

The first tool (identifying project leaders) will be provided to a randomly chosen 2/3^{rds} of the communities in our original study sample. It will be implemented as follows: during a focus group discussion with community leaders, a local facilitator will advertise the project challenge competition, explain how the process works, and detail what skills are needed to submit a strong proposal (i.e., write a project description, develop an itemized budget, and meet project submission deadlines). The facilitator will then ask the group to deliberate and identify five individuals, explicitly excluding the village headman, who are most likely to possess these skills. The identified individuals will complete a short management test in private; the test includes questions assessing basic math and writing skills, and testing knowledge about the cost of materials commonly used in small scale infrastructure projects in their area. The facilitator will grade the test on site. The facilitator will then reconvene the focus group to share the results of which individual had the highest score on the test, and will encourage the group to endorse this person to lead the project proposal development and submission process.

The second tool (basic training) will be offered to half of those communities that receive the first tool, in other words, a random 1/3rd of all communities. For these, the facilitator will announce the date and location of one-day management training sessions that the local governments are providing as part of the project challenge competition. The trainings will cover topics related to the identification of local community development needs and how to design projects to effectively address them; budgeting practices; and time management. Travel costs to the training will be covered for the individual selected to be the project proposal leader, and the facilitator will encourage that person to attend one of the trainings.

The remaining 1/3rd of study villages will serve as controls. The facilitator will announce the project challenge competition but will neither provide management tests scores nor advertise the training sessions. We expect many of these project proposal efforts to be directed by the village headman, which is the status quo in many communities in rural Sierra Leone.

We would like to know your views about what percentage of villages in the 6 different treatment assistance categories you think will actually submit a project proposal to the local government as part of the project challenge competition. As a reference point, we studied a voucher program in 2009 that subsidized the cost of construction materials by 34% (which could be for public or private use). 53% of communities in this sample took advantage of the program.

Types of assistance that different groups of communities received		**Percent that will submit a proposal (0 to 100%)**
CONTROL villages that **did NOT participate in GoBifo CDD (2005-09)**		
	Control communities that receive BOTH the manager selection tool AND training	|_____|%
	Control communities that receive the first manager selection tool but NO training	|_____|%
	Control communities that receive neither new tool (status quo)	|_____|%
TREATED villages that **DID participate in GoBifo CDD (2005-09)**		
	GoBifo communities that receive BOTH the manager selection tool AND training	|_____|%
	GoBifo communities that receive the first manager selection tool but NO training	|_____|%
	GoBifo communities that receive neither new tool (status quo)	|_____|%

You made 6 forecasts above about the new experiment. How many of these forecasts do you think will fall within 10 percentage points of the true effect size? _____(out of 6)

Appendix C: Additional Specifications

Table A2: Lower Imputation Bound, Treatment Effects on Grants Competition Performance

	Proposal Score (index)	Technical Score	Expert Score	Gov't Score
	(1)	(2)	(3)	(4)
Panel A: Technocratic Selection versus CDD				
Technocratic Selection	0.362**	0.465**	0.354**	0.267
	(0.168)	(0.191)	(0.172)	(0.179)
CDD	0.132	0.073	0.125	0.199
	(0.175)	(0.192)	(0.188)	(0.184)
Technocratic Selection * CDD	0.051	-0.025	0.173	0.006
	(0.221)	(0.247)	(0.232)	(0.236)
Observations	236	236	236	236
F -statistic (on TS and TS*CDD)	6.24	6.68	7.53	2.64
p -value	0.002	0.002	0.001	0.073
Panel B: Technocratic Selection and Managerial Training				
Technocratic Selection	0.252*	0.352**	0.245*	0.158
	(0.148)	(0.166)	(0.147)	(0.160)
Training	0.366**	0.311*	0.469***	0.319*
	(0.148)	(0.174)	(0.142)	(0.164)
F -statistic (on TS and TR)	10.39	9.23	13.99	5.17
p -value	<0.001	<0.001	<0.001	0.006
Observations	236	236	236	236

*Notes: i) significance levels indicated by * p < 0.10, ** p <0.05, *** p < 0.01; ii) robust standard errors; iii) specifications in Panel A pool the technocratic selection and training arms together (see Appendix Table A4 for full interaction model) and include strata for geographic ward and two balancing variables (distance to road and community size) from the original randomization; iv) specifications in Panel B include the two balancing variables and strata for ward crossed with CDD assignment; v) outcomes in columns 2 to 4 are mean effects indices, expressed in standard deviation units, standardized with respect to the mean and standard deviation of control Arm 1 (Arms 1 and 4) in Figure 1 for Panel A (B) (see Kling, Liebman and Katz 2007); vi) missing scores for the 4 non-submitting communities are imputed at the lowest observed score in the data; vii) outcome in column 1 is an equally weighted index of those in columns 2 to 4; vii) Training term in Panel B captures the additional effect of training beyond that of technocratic selection; ix) the F-statistic and associated p-value evaluate the hypothesis that the listed terms are jointly equal to zero; and x) the sample for all specifications includes all communities in Figure 1.*

Table A3: Upper Imputation Bound, Treatment Effects on Grants Competition Performance

	Proposal Score (index)	Technical Score	Expert Score	Gov't Score
	(1)	(2)	(3)	(4)
Panel A: Technocratic Selection versus CDD				
Technocratic Selection	0.366**	0.486**	0.348**	0.265
	(0.169)	(0.196)	(0.172)	(0.181)
CDD	0.001	-0.072	0.003	0.071
	(0.183)	(0.207)	(0.192)	(0.191)
Technocratic Selection * CDD	0.123	0.053	0.240	0.077
	(0.223)	(0.254)	(0.232)	(0.239)
Observations	236	236	236	236
F -statistic (on TS and TS*CDD)	7.71	8.34	8.82	3.40
p -value	0.001	<0.001	<0.001	0.035
Panel B: Technocratic Selection and Managerial Training				
Technocratic Selection	0.324**	0.438***	0.306**	0.228
	(0.141)	(0.158)	(0.143)	(0.154)
Training	0.309**	0.252	0.415***	0.260*
	(0.134)	(0.155)	(0.132)	(0.156)
F -statistic (on TS and TR)	11.34	10.71	14.63	5.27
p -value	<0.001	<0.001	<0.001	0.006
Observations	236	236	236	236

*Notes: i) significance levels indicated by * p < 0.10, ** p <0.05, *** p < 0.01; ii) robust standard errors; iii) specifications in Panel A pool the technocratic selection and training arms together (see Appendix Table A4 for full interaction model) and include strata for geographic ward and two balancing variables (distance to road and community size) from the original randomization; iv) specifications in Panel B include the two balancing variables and strata for ward crossed with CDD assignment; v) outcomes in columns 2 to 4 are mean effects indices, expressed in standard deviation units, standardized with respect to the mean and standard deviation of control Arm 1 (Arms 1 and 4) in Figure 1 for Panel A (B) (see Kling, Liebman and Katz 2007); vi) missing scores for the 4 non-submitting communities are imputed at the highest observed score in the data; vii) outcome in column 1 is an equally weighted index of those in columns 2 to 4; vii) Training term in Panel B captures the additional effect of training beyond that of technocratic selection; ix) the F-statistic and associated p-value evaluate the hypothesis that the listed terms are jointly equal to zero; and x) the sample for all specifications includes all communities in Figure 1.*

Table A4: Full Interaction Model

	Proposal Score (index)	Technical Score	Expert Score	Gov't Score	Won a Grant
	(1)	(2)	(3)	(4)	(5)
Technocratic Selection	0.312	0.430*	0.289	0.217	0.101
	(0.194)	(0.231)	(0.199)	(0.209)	(0.066)
Training	0.162	0.185	0.165	0.138	0.003
	(0.197)	(0.234)	(0.194)	(0.218)	(0.078)
CDD	0.057	-0.018	0.056	0.132	0.049
	(0.182)	(0.207)	(0.193)	(0.191)	(0.047)
Technocratic Selection * CDD	-0.076	-0.076	-0.058	-0.094	-0.070
	(0.267)	(0.307)	(0.273)	(0.287)	(0.088)
Training * CDD	0.349	0.192	0.564**	0.290	-0.033
	(0.255)	(0.308)	(0.253)	(0.292)	(0.099)
F -statistic (on TS, TR and interactions)	8.33	5.83	11.88	3.42	1.16
p -value	<0.001	<0.001	<0.001	0.010	0.331
F -statistic (on CDD and interactions)	1.44	0.14	4.07	1.11	0.63
p -value	0.233	0.939	0.008	0.345	0.597
F -statistic (on TS, TR, CDD and interactions)	8.05	4.68	11.93	3.45	1.01
p -value	<0.001	<0.001	<0.001	0.005	0.414
Observations	236	236	236	236	236

*Notes: i) significance levels indicated by * p < 0.10, ** p <0.05, *** p < 0.01; ii) specifications include strata for geographic ward and two balancing variables (distance to road and community size) from the original randomization; iii) robust standard errors; iv) outcomes coded to treatment arm mean for communities that did not submit a proposal in columns 2 to 5; v) outcomes in columns 2 to 4 are mean effects indices, expressed in standard deviation units, standardized with respect to the mean and standard deviation of Arm 1 in Figure 1 (see Kling, Liebman and Katz 2007); vi) outcomes in column 1 are an equally weighted index of those in columns 2 to 4; vii) outcomes in column 5 are expressed in proportions; viii) the F-statistic and associated p-value evaluate the hypothesis that the listed terms are jointly equal to zero; and ix) the sample for all specifications includes all communities in Figure 1 (Arms 1 to 6).*

Table A5: Two-way Comparison of Technocratic Selection and CDD

	Proposal Score (index)	Technical Score	Expert Score	Gov't Score	Won a Grant
	(1)	(2)	(3)	(4)	(5)
Technocratic Selection	0.444***	0.534***	0.487***	0.312**	0.059
	(0.113)	(0.129)	(0.118)	(0.120)	(0.036)
CDD Treatment	0.123	-0.004	0.208*	0.167	-0.008
	(0.105)	(0.123)	(0.109)	(0.116)	(0.037)
Observations	236	236	236	236	236
F -statistic (on TS and CDD)	8.74	8.52	10.47	4.45	1.31
p -value	<0.001	0.003	<0.001	0.013	0.272
F -statistic (TS = CDD)	4.10	8.94	2.92	0.75	1.36
p -value	0.044	0.003	0.089	0.388	0.244

*Notes: i) significance levels indicated by * p < 0.10, ** p <0.05, *** p < 0.01; ii) specifications include strata for geographic ward and two balancing variables (distance to road and community size) from the original randomization; iii) robust standard errors; iv) outcomes coded to treatment arm mean for communities that did not submit a proposal in columns 2 to 5; v) outcomes in columns 2 to 4 are mean effects indices, expressed in standard deviation units, standardized with respect to the mean and standard deviation of Arm 1 in Figure 1 (see Kling, Liebman and Katz 2007); vi) outcomes in column 1 are an equally weighted index of those in columns 2 to 4; vii) the F-statistics and associated p-values evaluate the hypothesis that the listed terms are jointly equal to zero, or equal to eachother; viii) outcomes in column 5 are expressed in proportions.*

Table A6: Technocratic Selection Effects for Simulated Winning Thresholds

	Winner, actual	Winner, 25th percentile	Winner, 50th percentile	Winner, 75th percentile
	(1)	(2)	(3)	(4)
Technocratic Selection	0.067	0.101	0.113	0.088
	(0.044)	(0.069)	(0.075)	(0.066)
Training	-0.026	0.090	0.167**	0.051
	(0.048)	(0.064)	(0.076)	(0.072)
F -statistic (on TS and TR)	1.28	3.96	6.82	2.25
p -value	0.281	0.021	0.001	0.108
Implied number of grants	20	178	120	61
Observations	236	236	236	236

*Notes: i) significance levels indicated by * p < 0.10, ** p <0.05, *** p < 0.01; ii) robust standard errors; iii) specifications include strata for geographic ward crossed with CDD assignment; iv) outcomes in column (2)-(4) are binary indicator for winning a grant at percentiles of the government proposal score distribution; and v) the F-statistic and associated p-value evaluate the hypothesis that the listed terms are jointly equal to zero.*

Table A7: Delegation to Trained versus Untrained Technocrats

Training Effect on Delegation	Technocratic Selection (arms 2, 5)	Training (arms 3, 6)	*p*- value on difference
Proportion where chiefly authorities chose the project	0.43	0.48	0.59
Proportion where chiefly authorities wrote the description	0.20	0.31	0.13
Proportion where chiefly authorities did the budget	0.19	0.26	0.33
Proportion where chiefly authorities set the timeline	0.20	0.32	0.09
Observations	148		

Notes: outcomes capture the propotion of management decisions that were made by the village headman or other chiefly authorities in the community and compares technocrates with and without training.

Table A8: Management Training and "Teaching to the Test"

	Panel A: "Copycat" measures				Panel B: Performance spillover measures		
	References sustainability	References multiple bids	References skills needed	Index	Says who will benefit	Says where items bought	Index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Technocratic Selection	-0.100	0.231	0.329*	0.153	0.163	0.129	0.146
	(0.148)	(0.218)	(0.177)	(0.103)	(0.147)	(0.158)	(0.108)
Training	0.362**	-0.231	0.127	0.086	0.125	-0.320**	-0.097
	(0.179)	(0.225)	(0.198)	(0.114)	(0.127)	(0.153)	(0.103)
Constant	0.028	0.551	-0.032	0.182	0.192	-0.042	0.075
	(0.354)	(0.704)	(0.370)	(0.288)	(0.213)	(0.290)	(0.132)
Observations	236	236	236	236	236	236	236

*Notes i) significance levels indicated by * p < 0.10, ** p <0.05, *** p < 0.01; ii) specifications include fixed effects for geographic ward crossed with CDD assignment; iii) Panel A looks for evidence of "teaching to the test" by seeing whether trainees mechanically include reference in their proposals to topics covered by the training but not asked for on the application (e.g. the training emphasized the value of seeking multiple bids from contractors during project construction, a good practice for winners to use during implementation but not something that the application required, and column 2 shows that trainees were no more likely to include extraneous reference to it in their proposals); iv) Panel B takes the converse approach and evaluates whether the training had performance spillover effects on application questions that were not addressed in the training (e.g. the application asked for an explanation of who would benefit from the project, a topic not discussed during the training, and column 5 shows that trainees were no more conscientious in including explanation of who benefits in their proposal); and v) outcomes in columns 4 and 7 are summary indices for the multiple measures in each panel.*

Table A9: 2018 Infrastructure Assessment of Grant Competition Winners by Treatment Assignments

		Technocratic Selection Experiment				CDD Experiment			
	Mean, full sample	Mean, technocrats	Mean, Status Quo Chiefs	Difference	*p*-value	CDD treatment	CDD control	Difference	*p*-value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Is the infrastructure present and functional?	0.70	0.63	1.00	-0.38	0.16	0.89	0.55	0.34	0.11
Quality of construction (1=poor, 10=excellent)	6.80	6.56	7.75	-1.19	0.26	7.00	6.64	0.36	0.67
Total community financial contributions (US$)	218.3	173.8	396.5	-222.7	0.14	233.6	205.9	27.7	0.83
Infrastructure is located near chief's compound	0.40	0.38	0.50	-0.12	0.67	0.33	0.45	-0.12	0.61
Observations	20	16	4			9	11		

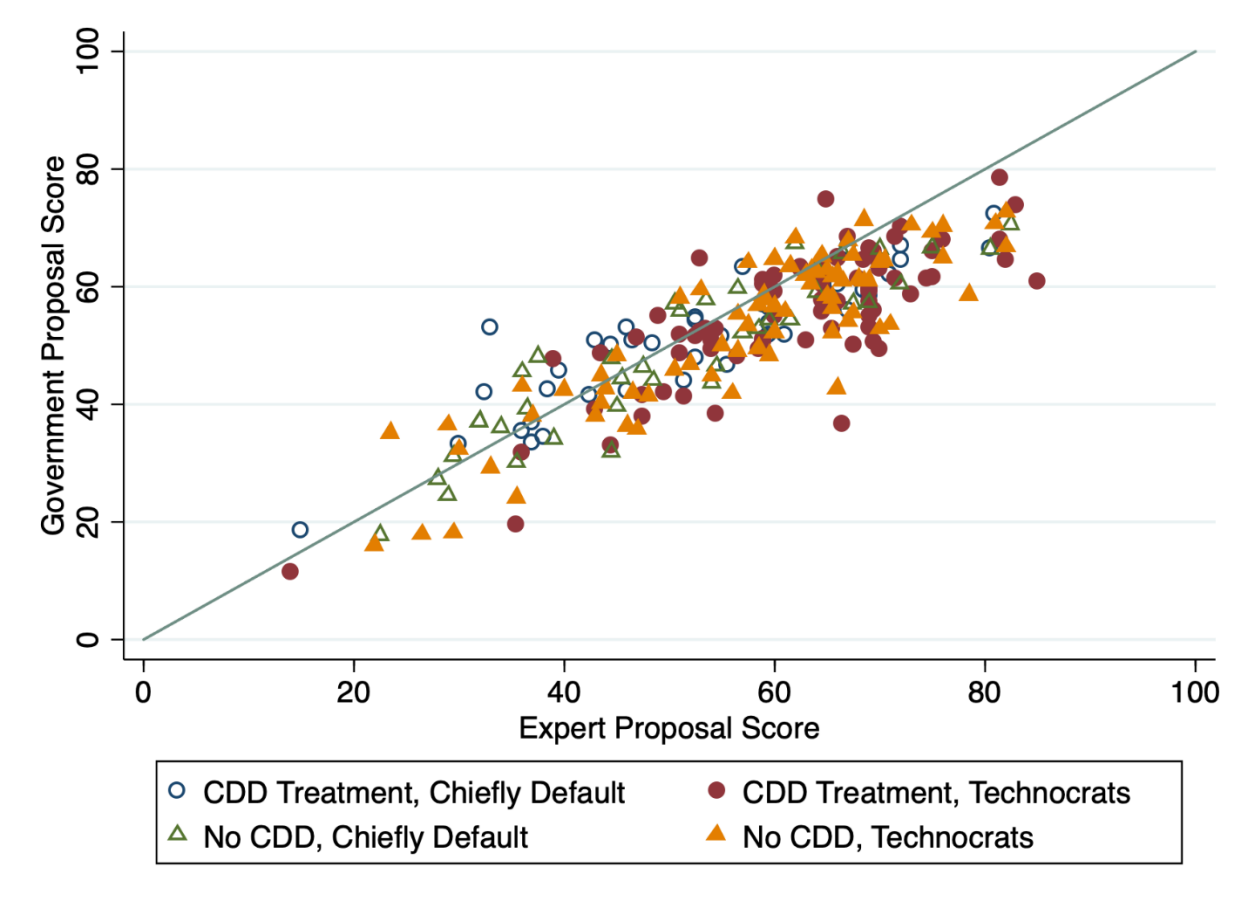
*Notes i) significance levels indicated by * p < 0.10, ** p <0.05, *** p < 0.01; ii) data is from the July 2018 field inspection of infrastructure projects that won an implementation grant from the government competition; and iii) estimates displayed are from two-sided t-tests for each of the two distinct experimental assignments.*

Table A10: Text Analysis of Proposal Content Across Treatment Assignment

	Proposal mentions inclusiveness terms	Proposal mentions community institutions	Community Center project	Education project	Water project	Other project
	(1)	(2)	(3)	(4)	(5)	(6)
CDD	0.063	0.024	-0.116*	0.050	0.051*	0.014
	(0.101)	(0.062)	(0.060)	(0.037)	(0.030)	(0.064)
Technocratic Selection	-0.040	0.134**	-0.043	0.024	-0.038	0.058
	(0.109)	(0.068)	(0.064)	(0.038)	(0.035)	(0.069)
Observations	236	236	232	232	232	232

*Notes i) significance levels indicated by * p < 0.10, ** p <0.05, *** p < 0.01. ii) specifications include strata for geographic ward and two balancing variables (distance to road and community size) from the randomization; and iii) robust standard errors.*

Figure A1: Distribution of Government and Practitioner Scores by Treatment Assignment



Notes: This figures plots the distribution of proposal scores given by the district government officials to allocate the infrastructure grants (Y-axis) against the scores given by unaffiliated development practitioners using the same scoring guidelines (X-axis). Higher scores indicate higher quality proposals. Both sets of raters were blinded to the name of the submitting communities. Each dot represents a proposal submitted by a particular community, where triangles indicate CDD treatment status, circles indicate CDD control status, shaded in shapes indicate assignment to technocratic selection and hollow shapes indicate assignment to the chiefly control default condition.

Table A11: Treatment Effects on Grants Competition Performance Controlling for Project Type

	Proposal Score (index)	Technical Score	Expert Score	Gov't Score	Won a Grant
	(1)	(2)	(3)	(4)	(5)
Technocratic Selection	0.378**	0.511***	0.355**	0.268	0.098*
	(0.166)	(0.196)	(0.170)	(0.181)	(0.051)
CDD	0.084	0.027	0.072	0.154	0.058
	(0.183)	(0.209)	(0.194)	(0.194)	(0.052)
Technocratic Selection * CDD	0.114	0.013	0.260	0.068	-0.090
	(0.225)	(0.260)	(0.235)	(0.244)	(0.072)
Community Center Project	-0.059	-0.128	0.020	-0.069	0.006
	(0.122)	(0.141)	(0.130)	(0.141)	(0.049)
School/Education Project	-0.475**	-0.499*	-0.498**	-0.429*	-0.054
	(0.224)	(0.260)	(0.218)	(0.239)	(0.054)
Water/Sanitation Project	-0.316	-0.546*	-0.114	-0.287	-0.122**
	(0.263)	(0.321)	(0.285)	(0.252)	(0.058)
Observations	232	232	232	232	232
F -stat (on TS and TS*CDD)	7.69	8.02	9.01	3.16	1.89
p- value	<0.001	<0.001	<0.001	0.044	0.154

*Notes: i) significance levels indicated by * p < 0.10, ** p <0.05, *** p < 0.01; ii) robust standard errors; iii) project type fixed effects denote the sectoral type of project proposed by the community in the grants competition application; iv) specifications pool the technocratic selection and training arms together and include strata for geographic ward and two balancing variables (distance to road and community size) from the original randomization; v) outcomes in columns 2 to 4 are mean effects indices, expressed in standard deviation units, standardized with respect to the mean and standard deviation of control arm 1 (arms 1 and 4) in Figure 1 for Panel A (B) (see Kling, Liebman and Katz 2007); vi) missing scores for the 4 non-submitting communities are imputed at the respective treatment arm mean; vii) outcome in column 1 is an equally weighted index of those in columns 2 to 4; viii) outcome in column 5 is a binary indicator; ix) the F-statistic and associated p-value evaluate the hypothesis that the listed terms are jointly equal to zero; and x) sample includes all communities in Figure 1.*

Table A12: Long-run CDD Treatment Effects on Exact Panel Outcomes

	Treatment effect 2016	Naïve *p*-value	FDR *q*-value	Treatment effect 2009	Change over time
	(1)	(2)	(3)	(4)	(1) - (4)
Panel A: Institutions Family					
All outcomes in family (*N*=56)	0.064**	0.010	0.006	0.086***	0.037
	(0.027)			(0.030)	(0.028)
Collective action	0.104*	0.050	0.234	0.072	0.086
	(0.053)			(0.046)	(0.061)
Inclusion	0.034	0.351	0.539	0.084*	0.031
	(0.036)			(0.049)	(0.044)
Local authority	-0.032	0.573	0.632	0.110	-0.088
	(0.056)			(0.068)	(0.070)
Trust	0.107*	0.065	0.234	0.032	0.064
	(0.057)			(0.049)	(0.081)
Groups and networks	0.149**	0.038	0.234	0.056	0.121
	(0.071)			(0.045)	(0.074)
Access to information	-0.036	0.590	0.632	0.150**	-0.075
	(0.067)			(0.072)	(0.072)
Participation in local governance	0.079	0.191	0.348	0.256***	-0.011
	(0.060)			(0.058)	(0.065)
Crime and conflict	-0.002	0.971	0.76	0.088	-0.012
	(0.063)			(0.062)	(0.074)
Political and social attitudes	0.154	0.216	0.348	-0.020	0.113
	(0.124)			(0.080)	(0.126)
Panel B: Infrastructure Family					
All outcomes in family (*N* = 29)	0.208***	<0.001	0.001	0.352***	-0.094***
	(0.041)			(0.035)	(0.036)
Project implementation	0.287***	<0.001	<0.001	0.875***	-0.450***
	(0.075)			(0.062)	(0.081)
Local public goods	0.228***	<0.001	<0.001	0.210***	0.024
	(0.046)			(0.041)	(0.041)
Economic welfare	0.240***	<0.001	<0.001	0.606***	-0.136**
	(0.056)			(0.061)	(0.062)
Observations	236			236	236

*Note: i) significance levels based on naive p-values and indicated by * p<0.10, ** p<0.05, ***p<0.01. ii) specifications include strata for geographic ward and two balancing variables (distance to road and community size) from the randomization; iii) robust standard errors; iv) all estimates are for hypothesis-level mean effects indices that equally weight component measures and are expressed in standard deviation units (see Kling, Liebman and Katz 2007); v) outcomes limited to those that were collected in the exact same fashion in both 2009 and 2016 survey rounds; and vi) 2009 data sourced from Casey et al (2012).*

Appendix D: CDD Effects on Community Response to Ebola

Table A13: CDD Treatment Effects on Ebola Responsiveness

Outcome	Mean, controls	Treatment effect	Standard error	*p*-value
Mean Effects Index (all 13 indicators)	0.000	0.042	0.038	0.27
Community had an Ebola task force during the Ebola crisis	0.661	0.077	0.052	0.144
Correctly answers "No" to "Can Ebola spread through air?"	0.856	-0.005	0.040	0.896
Correctly answers "21" to "How many DAYS can it take for the first to symptoms arise?"	0.669	0.014	0.051	0.791
Total (of 11 possible) correct answers to questions about how one can get Ebola	5.220	0.006	0.187	0.974
Knows correct Ebola hotline number	1.000	0.000	0.000	.
Community created bye-laws in relation to Ebola	0.907	0.042**	0.019	0.029
Total (of 10 possible) correct answers regarding how to protect yourself against Ebola	4.975	-0.051	0.201	0.801
Correctly answers "No" to "Drinking salt water can help cure Ebola?"	0.958	0.030	0.019	0.112
Correctly answers "No" to "Drinking chloring can help cure Ebola?"	1.000	-0.009	0.009	0.319
Communities are more likely to go to formal health facilities (nurse, clinic)	0.924	0.014	0.030	0.631
Communities are more likely to go to formal health facilities for Ebola (nurse, clinic)	0.915	0.000	0.034	0.995
Correctly answers "No" to "Can someone spread Ebola even before they show symptoms?	0.695	0.030	0.052	0.564
Total correct answers (of 14 possible) regarding symptoms of Ebola	7.263	-0.230	0.232	0.323
Observations	236			

*Note: i) significance levels based on naive p-values and indicated by * p<0.10, ** p<0.05, ***p<0.01. ii) specifications include strata for geographic ward and two balancing variables (distance to road and community size) from the randomization; iii) robust standard errors; and iv) this table includes 13 of 15 pre-specified primary outcomes in our PAP, excluding 2 outcomes that are observed for fewer than 20 communities in the data.*

Appendix E: Raw Results for Long-run CDD Effects on Individual Outcome Measures

Appendix Table A14: Raw Results for CDD Effects on Individual Outcomes

Row	Variable	Hypothesis	Mean, controls	Treatment effect	Standard error	Naïve *p*-value	N
H1: Implementation							
1	Does this community have a bank account?	H1, H3	0.042	0.240	0.040	0.000	236
2	Average score of all test takers	H1	41.681	1.145	1.500	0.445	235
3	Does this community have a Village or Community Development Committee (VDC or CDC)?	H1, H4, H10	0.432	0.173	0.057	0.003	236
4	Does this community have a village development plan (i.e. an agreed plan with specific priorities for what the community will do for its own development over the next few years)?	H1, H10	0.492	0.003	0.057	0.955	236
5	Has this community been visited by a Local Councillor in the past one year?	H1, H9	0.263	-0.074	0.046	0.109	236
6	Has this community been visited by a Ward Development Committee member in the past year?	H1, H9	0.102	0.019	0.035	0.579	236
H2: GoBifo improves the quality of local public services infrastructure.							
7	Ask the community: when was the last time this community brushed this foot path?	H2, H4	-35.224	1.123	4.707	0.811	234
8	Does the community have a court barrie and is it functional?	H2	0.102	0.218	0.040	0.000	236
9	Does the community have a community center and is it functional?	H2	0.068	0.060	0.038	0.111	236
10	Does the community have a drying floor and is it functional?	H2	0.178	0.127	0.051	0.013	236
11	Does the community have a grain store and is it functional?	H2	0.119	0.198	0.051	0.000	236
12	Does the community have a latrine and is it functional?	H2	0.076	0.029	0.036	0.412	236
13	Does the community have a market and is it functional?	H2	0.000	0.025	0.013	0.064	236
14	Does the community have a palava hut and is it functional?	H2	0.042	0.019	0.028	0.487	236
15	Does the community have a public health unit and is it functional?	H2	0.110	-0.022	0.038	0.565	236
16	Does the community have a primary school and is it functional?	H2	0.466	0.125	0.058	0.030	236
17	Does the community have any wells (mechanical or bucket) and are any of them functional?	H2	0.661	0.000	0.057	0.997	236
18	Do any of the local sports teams have uniforms / vests?	H2	0.153	0.003	0.046	0.946	236
19	Does the community have a football / sports field and is it functional?	H2	0.619	0.160	0.054	0.003	236
20	Does the community have a traditional birth attendant (TBA) house and is it	H2	0.025	0.124	0.032	0.000	236

Appendix Table A14: Raw Results for CDD Effects on Individual Outcomes (continued)

Row	Variable	Hypothesis	Mean, controls	Treatment effect	Standard error	Naïve *p*-value	N
21	Ask to be taken to the nearest bush path. This should be a foot path (not a road for cars) that the community uses the most. Walk 100 steps down the path (i.e. look at the middle, not the start of the path). In your own opinion, how bushy is the path? [Answer indexed from 0 "very bushy" to 1 "very clear"]	H2, H4	2.653	-0.049	0.110	0.658	236
22	Since January 2006, has this community taken a project proposal to an external funder—like local government or NGO—for support? *Note that the community should have been the ones initiating the request.*	H2, H4	0.246	0.048	0.054	0.370	236
23	Does this community have a seed bank (i.e. where people can borrow rice or groundnuts to plant and repay after harvest)?	H2	0.085	0.049	0.040	0.225	236
H3: GoBifo improves general economic welfare							
24	Supervisor assessment that community is "much better off" or "a little better off" than other communities he/she has been to in this area	H3	0.364	0.091	0.058	0.114	236
25	When was the last time an outsider trader came to this village to buy agricultural or non-agricultural goods? (date - date of interview)	H3	-12.178	3.468	4.820	0.472	236
26	[From supervisor tour of community] Have you seen anybody selling packaged goods (cigarettes, crackers, etc) in this village today from their own home (i.e. not out of a store)?	H3	0.881	-0.015	0.040	0.705	236
27	Number of goods out of 10 common items (bread, soap, garri, country cloth/garra tie-dye, eggs/chickens, sheep/goats, palm oil/nut oil, coal, carpenter for hire/shop, tailor/dressmaker, blacksmith for hire/shop) that you can buy in this community today	H3	5.619	0.403	0.247	0.103	236
28	How many people have started a new business (even if it is small or informal) in this community in the past 2 years (since October 2007)? [Record name, type of business and year started]	H3	6.297	0.627	0.500	0.210	236
29	How many houses and small shops (including tables, boxes and kiosks) are selling packaged goods (like cigarettes, biscuits, etc) inside this community today?	H3	3.737	0.626	0.343	0.068	236
30	In the past 2 years (since October 2007), have you participated in any skills training (bookkeeping, soap-making), adult literacy (learn book) or vocation education courses (carpentry, etc.)?	H3	2.831	0.270	0.629	0.667	236
H4: GoBifo increases collective action and contribution to local public goods.							
31	Does this community have any communal farms?	H4	0.144	0.087	0.049	0.073	236
32	Does the primary school that children in the community attend have community	H4	0.746	0.066	0.049	0.179	236
33	Average quality of proposal as assessed by experts	H4	55.309	3.247	1.807	0.072	232
34	Do any people from different households here come together to sell agricultural goods or other petty trading as a group to markets outside of this village (i.e. heap the goods together and send one person to sell; NOT every person totes their own load)?	H4, H7, H8	0.347	-0.046	0.053	0.390	236

Appendix Table A14: Raw Results for CDD Effects on Individual Outcomes (continued)

Row	Variable	Hypothesis	Mean, controls	Treatment effect	Standard error	Naïve *p*-value	N
35	Average quality of proposal as assessed by policy makers	H4	51.262	2.461	1.591	0.122	232
36	Average completeness of proposal	H4	10.026	-0.013	0.283	0.964	232
37	Whether the proposal is among the top 20 and a winner (as ranked by the Gobifo staff ar	H4	0.096	-0.011	0.037	0.767	232
38	Do any disabled people hold leadership positions in this community (like member of VDC, youth leaders, headman, women's leader, secret society head)?	H5	0.144	0.033	0.048	0.499	236
H5: GoBifo increases inclusion and participation in community planning and implementation, especially for poor and vulnerable groups; GoBifo norms spill over							
39	Did any disabled people (blind, polio, amputee, wheelchair, etc.) attend the last community meeting?	H5	0.398	0.102	0.063	0.103	236
40	In the past one year, have you attended any community meetings?	H5	-28.644	7.510	7.084	0.289	236
41	Enumerator record of total women (18+ years) present at gift choice meeting (field activity #1)	H5	2.449	-0.031	0.141	0.828	236
42	Enumerator record of total youths (18-35 years) present at gift choice meeting (field activity #1)	H5	2.288	-0.209	0.193	0.280	236
43	Did anyone take minutes (written record of what was said) at the most recent community meeting?	H5	0.220	0.075	0.056	0.181	236
44	Less concentrated deliberation in manager selection	H5	2.892	0.023	0.090	0.798	231
45	Less concentrated deliberation in manager selection	H5	1.416	0.013	0.057	0.813	192
46	Enumerator account of how democratically the group evenutally came to a decision about who the potential project managers ranging from 5 = open discussion followed by group vote to 1 = chief and/or elders decide without other input	H5	3.364	-0.002	0.094	0.982	235
47	Time of deliberation of manager selection process	H5	32.486	53.665	27.838	0.054	210
48	Enumerator record of total public speakers durings selection of potential project managers	H5	43.429	-2.772	2.584	0.283	213
49	Did a vote occur during the project leader nomination discussion	H5, H6	1.929	0.023	0.032	0.463	171
50	Record of total women (18+ years) in "important people" focus group list	H5	13.264	-0.570	1.176	0.628	216
51	Enumerator account of how actively women participated in the deliberation on the selection of potential project managers compared to men, ranging from 5 = no difference between women and men to 1 = women not active at all compared to men	H5	2.799	-0.122	0.132	0.356	232
52	Record of total youth (18-35 years) in "important people" focus group list	H5	6.009	-0.402	0.391	0.303	229
53	Enumerator account of how actively youth participated in the deliberation on the selecti	H5	3.035	0.173	0.153	0.259	229
54	Has this community had any problems with financial mismanagement/corruption in the past 2 years (since November 2014)?	H5	0.839	-0.020	0.044	0.656	236

Appendix Table A14: Raw Results for CDD Effects on Individual Outcomes (continued)

Row	Variable	Hypothesis	Mean, controls	Treatment effect	Standard error	Naïve *p*-value	N
H6: GoBifo changes local systems of authority, including the roles and public perception of traditional leaders (chiefs) versus elected local government.							
55	How old is the current (or acting) village chief/ Headman?	H6	-59.301	-0.974	1.830	0.595	228
56	Enumerator reports on whether "chief decided" project leader nominations	H6	0.873	-0.050	0.043	0.241	235
57	Relative view of "do people in this community believe" Local Councilors as opposed to Chiefdom officials	H6	-0.119	-0.021	0.052	0.683	236
H7: GoBifo increases trust							
58	Are you a member of any credit or savings (osusu) groups?	H7, H8	2.432	0.476	0.285	0.095	236
59	In general, do people in this community believe the central government officials or do they think you need to be careful when dealing with them?	H7	0.314	0.013	0.051	0.794	236
60	In general, do people in this community believe chiefdom officials or do you have to be careful when dealing with them?	H7	0.195	0.053	0.048	0.272	236
61	In general, do people in this community believe Local Councillors or do you have to be careful when dealing with them?	H7	0.076	0.032	0.037	0.391	236
62	In general, do people in this community believe NGOs / donor projects or do you have to be careful when dealing with them?	H7	0.500	0.168	0.057	0.003	236
63	In general, do people in this community believe people from outside you own village / town / neighborhood or do you have to be careful when dealing with them?	H7	0.127	0.088	0.047	0.062	236
64	In general, do people in this community believe people from you own village / town / neighborhood or do you have to be careful when dealing with them?	H7	0.703	-0.069	0.057	0.224	236
H8: Gobifo builds and strengthens community groups and networks							
65	Are there any fishing groups / cooperatives in this community?	H8	0.246	0.037	0.042	0.380	236
66	How many active school PTA groups are there in this village?	H8	4.076	0.719	1.208	0.552	236
67	How many active religious groups (not just going to church/mosque) are there in this village?	H8	4.102	1.721	2.019	0.394	236
68	How many active groups for saving for special events (weddings, funerals) are there in this village?	H8	0.517	0.164	0.116	0.156	236
69	How many active seed multiplication groups are there in this village	H8	0.254	0.853	0.485	0.079	236
70	How many active social clubs are there in this village?	H8	1.441	0.183	0.164	0.264	236
71	How many active women's groups (general) are there in this village?	H8	0.983	-0.039	0.124	0.749	236
72	How many active youth groups (general) are there in this village?	H8	1.212	0.013	0.110	0.907	236

Appendix Table A14: Raw Results for CDD Effects on Individual Outcomes (continued)

Row	Variable	Hypothesis	Mean, controls	Treatment effect	Standard error	Naïve *p*-value	N
H9: GoBifo increases access to information about local governance							
73	Supervisor assessment of whether there are any of the following items--awareness campaigns, financial information, development plan, minutes from any meetings, government policies, election information--visible anywhere around the village (i.e. on a notice board, school, clinic, shop, etc.)?	H9	0.117	0.005	0.018	0.805	236
74	Has this community been visited by the Paramount Chief in the past year?	H9	0.127	-0.023	0.040	0.561	236
H10: GoBifo increases public participation in local governance							
75	Did anyone in this community contest the party symbol in the 2008 local council elections?	H10	0.169	-0.006	0.044	0.899	236
76	Did anyone in this community stand for the most recent paramount chief elections?	H10	0.068	0.032	0.035	0.357	236
77	Did anyone in this community stand for the most recent section chief elections?	H10	0.280	0.016	0.057	0.777	236
78	Did anyone in this community stand for the most recent Ward Development Committee elections or get nominated for WDC?	H10	0.212	-0.011	0.048	0.813	236
H11: By increasing trust, GoBifo reduces crime and conflict in community.							
79	No conflict that respondent needed help from someone outside the household to resolve in the past one year	H11	-10.424	0.520	1.103	0.637	236
80	In the past 12 months, respondent has not been involved in any physical fighting	H11	-0.568	-0.124	0.270	0.646	236
81	In the past 12 months, no livestock, household items or money stolen from the	H11	-12.127	-1.406	1.267	0.267	236
82	During the last 12 months, respondent has not been a victim of witchcraft (juju)	H11	-1.441	0.441	0.351	0.208	236
H12: GoBifo changes political and social attitudes, making individuals more liberal towards women, more accepting of other ethnic groups and "strangers", and less tolerant of corruption and violence.							
83	Is the current (or acting) village chief/Headman a woman?	H12	0.034	-0.010	0.022	0.653	236
84	Is the current (or acting) village chief/Headman less than 35 years old?	H12	0.009	0.034	0.021	0.107	228

*Notes: i) significance levels (per comparison p-value) indicated by * p <0.10, ** p <0.05, *** p <0.01; ii) specification that includes fixed effects for the disctrict council wards (the unit of stratification) and the two balancing variables from the randomization (total households and distance to road) with robust standard errors; iii) "per comparison" p values are appropriate for a priori interest in an individual outcome*

Appendix F: Pre-analysis Plan

We include below the text of our pre-analysis plan with annotation to flag where the referenced specifications appear in the main text and appendix. The plan, with time stamps, can be found in the American Economic Association's registry for randomized control trials (https://www.socialscienceregistry.org/trials/1784), where detailed Excel sheets listing all outcome variables (referenced as "PAP Sheets 1, 2, 3 and 4") are also available for download.

Annotated to link pre-specifications to manuscript tables 25 September 2019

Pre-analysis Plan: Two Approaches to Community Development

10 March 2017

PIs: K. Casey, R. Glennerster, E. Miguel and M. Voors

Overview

This research project has four main components. The first evaluates the long run effects of a community driven development (CDD) program in Sierra Leone. The project devolved financial and implementation control over public services to communities, accompanied by intensive social facilitation. The second assesses a low cost technocratic alternative that identifies and supports high competence community members to take better advantage of development opportunities. It leverages local talent, addresses information barriers, and augments existing managerial capital with basic training in project management. A third component elicits expert beliefs about the efficacy of these two approaches and assesses their forecast levels and accuracy. A fourth line of inquiry examines whether participation in CDD affected community response to the Ebola crisis.

Registration timeline

We registered this study with the American Economic Association (AEA) Randomized Control Trial Registry on 16 November 2016. Our trial entry can be found here: http://www.socialscienceregistry.org/trials/1784. On 17 November 2016, we uploaded a data management plan that outlines who would have access to data when, and commits all PIs to not access any data with identifying information until after this PAP is lodged. Fieldwork commenced on 18 November 2016. Our Field Manager Angelica Eguiguren at IPA Sierra Leone was the only person who had access to the data at all times. She uploaded the data to a secure server and will invite the PIs to that dropbox as soon as the PAP is lodged. We lodged an email confirming PI adherence to the data management plan on 9 March 2017. We lodged this PAP on 10 March 2017. We have received IRB clearance from Stanford (#38846), the Government of Sierra Leone, Office of the Sierra Leone Ethics and Scientific Review Committee (3-11-2016, Wageningen (18-11-2016), Berkeley (2016099099) and MIT (#1612798296) for this trial.

Part I: Long run effects of CDD

Component Overview: Community Driven Development (CDD) is a participatory approach popular with foreign aid donors that involves communities directly in the financial management and implementation of local public goods. CDD has two main aims: i) improve the stock and quality of local public goods via the provision of block grants; and ii) democratize local decision-making via intensive social facilitation focused on the participation of marginalized groups.

In earlier work, we analyzed the medium run effects of the "GoBifo" CDD project in Sierra Leone (Casey, Glennerster and Miguel 2012).[1] GoBifo was implemented from 2005 to 2009 and provided roughly $5,000 in block grants and six months of dedicated social facilitation per community. The medium run study found substantial positive impacts on local public goods and economic activity, stronger links between the community and local government, and no evidence for more inclusive local decision-making.

[1] Casey K, Glennerster R, Miguel E (2012) Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan. Quarterly Journal of Economics 127 (4): 1755-1812.

During late 2016, we revisited the 236 communities in the original study to assess long term impacts. In the interim, 60 of the treatment communities received additional support from the GoBifo project. Specifically, these 60 communities received $1,300 for youth empowerment programs in 2010. We do not know how exactly the project management staff selected these 60 communities from the pool of 118 treatment communities, but it was not via random assignment.

Hypotheses: The 12 research hypotheses grouped into two families remain the same as those used in the earlier study.

- Family A of hardware outcomes: "GoBifo creates functional development committees" (H1); "Participation in GoBifo improves the quality of local public services infrastructure" (H2); and "Participation in GoBifo improves general economic welfare" (H3).

- Family B of software outcomes: "Participation in GoBifo increases collective action and contributions to local public goods" (H4); "GoBifo increases inclusion and participation in community planning and implementation, especially for poor and vulnerable groups; GoBifo norms spill over into other types of community decisions, making them more inclusive, transparent and accountable" (H5); "GoBifo changes local systems of authority, including the roles and public perception of traditional leaders (chiefs) versus elected local government" (H6);[2] "Participation in GoBifo increases trust" (H7); "Participation in GoBifo builds and strengthens community groups and networks" (H8); "Participation in GoBifo increases access to information about local governance" (H9); "GoBifo increases public participation in local governance" (H10); "By increasing trust, GoBifo reduces crime and conflict in the community" (H11); and "GoBifo changes political and social attitudes, making individuals more liberal towards women, more accepting of other ethnic groups and 'strangers', and less tolerant of corruption and violence" (H12).

Econometric Specifications: For Part I, the primary test of interest is evaluating long run effects of CDD at the family level. Our core specification evaluates treatment effects for Family A and B, using the following model:

$$Y_c^L = \beta_0 + \beta_1 T_c + X'_c \Gamma + W'_c \Pi + \varepsilon_c \quad (1)$$ In Table 4

where Y_c^L is the mean index for each family for community c in the 2016 survey round; T_c is the GoBifo treatment indicator; X_c contains two village-level balancing variables from the randomization process (distance from a road and total number of households); W_c is a fixed effect for geographic ward, the administrative level on which the randomization was stratified; and ε_c is the usual idiosyncratic error term. The parameter of interest is β_1, the average long run treatment effect. We will construct mean effects indices following Kling, Liebman and Katz (2007).[3]

To interpret these effects, we will test whether long run effects differ from the medium run effects in areas where the medium run effects were nonzero (Family A). Here we will test for decay using the following model:

$$Y_c^L - Y_c^M = \gamma_0 + \gamma_1 T_c + X'_c \Lambda + W'_c \Theta + \mu_c \quad (2)$$ In Table 4

[2] As before, that this is not an explicit objective of the GoBifo project leadership itself, but is a plausible research hypothesis.
[3] Kling, J., J. Lieberman and L. Katz (2007) Experimental Analysis of Neighborhood Effects, Econometrica, 75(1); 83–119

where the dependent variable is the difference in mean effects indices measured in the 2016 survey, Y_c^L, and 2009, Y_c^M. The coefficient of interest is γ_1, where $\gamma_1 < 0$ suggests that the treatment effect has dissipated over time for that hypothesis. A combination of failing to reject $\beta_1 = 0$ while rejecting $\gamma_1 \geq 0$ suggests that previously observed treatment effects have dissipated, while failing to reject $\beta_1 = 0$ and $\gamma_1 \geq 0$ presents a less conclusive middle ground that likely reflects greater noise in measuring long run outcomes and accompanying reductions in the power to detect treatment effects. Note that the exact set of outcomes varies between the 2009 and 2016 data collection rounds, so each index will incorporate the relevant outcomes for that particular survey round (see below).

The second test of interest is running Equations (1) and (2) at the hypothesis level where Equation (2) will again only be run for hypotheses with non-zero medium run effects.

Throughout our analysis, we will adjust for the fact that we are running more than one test on the same dataset by implementing false discovery rate (FDR) corrections. Research practice appears to be moving towards FDR and away from the more conservative familywise error rate (FWER) corrections where there are several tests of interest. Since our earlier paper used FWER corrections, we will also report them here to maintain consistency, but note that the preferred specifications use FDR. These adjustments run across the two families (Family A and Family B) or 12 hypotheses (H1 – H12) as relevant. See Benjamini, Krieger and Yekutieli (2006) and Anderson (2008).[4] For all tests, we will also report the "naïve" or "per comparison" *p*-value.

Our third test of interest highlights a few individual outcome measures from a new structured community activity (SCA). Here we will test for long run effects of GoBifo on the managerial capital of community members and the quality of proposals submitted to a project challenge competition run by the local District Councils (discussed in greater detail below). These outcomes measure whether the learning-by-doing experience of participating in GoBifo translates into long run differences in ability to act collectively and take advantage of development opportunities. We will test them as part of our larger research framework under H1 and H4, respectively, but also highlight them on their own as they capture an important channel through which GoBifo could lead to long run changes.

To further interpret the family- and hypothesis-level results, we will also estimate Equation (1) at the level of individual outcome (adjusting for FDR across all outcomes under a given hypothesis). Note that this reporting of all individual outcomes is for illustrative and interpretation purposes only.

Measurement and survey instruments: See ["SES - Endline 2016"]. The main data collection instrument for the long run effects closely follows the community modules used in the 2009 survey. This includes a focus group discussion with local leaders and enumerator physical inspection of community amenities and market activity. Where possible, we have included a community-level analogue of household level indicators included in the 2009 survey. In addition to economic and social outcomes, we include measures of institutional outcomes using the new project challenge SCA. These are captured in several instruments ["Managerial capital test", "Manager selection tally sheet enumerator A and B", "Submission survey", "Submission form", "Technical scoring", "Policy Scoring", "Expert Scoring"]. We did not repeat the household level survey due to budget constraints.

[4] Benjamini, Y., A. Krieger, and D. Yekutieli (2006) Adaptive Linear Step-Up Procedures That Control the False Discovery Rate, Biometrika, 93: 491–507. Anderson, M (2008) 'Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedaian, Perry Preschool, and Early Training Projects,' Journal of the American Statistical Association, 103 (484): 1481–1495.

Outcomes: See ["PAP, sheet 1"]. The table maps each individual outcome to the hypothesis of interest. To facilitate comparison to our earlier work, the first several columns of this table reproduce exactly those in the Appendix J: Raw Results from the supplementary materials to the 2012 QJE article. The list of outcomes has evolved in a few key ways. First, the present data collection uses only community modules and does not conduct household visits. Thus, all household level outcomes (indicated by "HH" in column K "2009 survey level") are omitted. Where possible, we have included a community-level analogue in the current survey (see column O "Additional question 2016"). Second, we exclude almost all conditional outcomes (i.e. those that are contingent on having a specific good in the community) that are only observed for a subset of villages. Third, as part of our new SCA, we designed measures that mirror some of the process-oriented 2009 SCA outcomes (e.g. unobtrusively counting the number of women who participate in a community decision).

The Casey et al (2012) paper included 334 outcomes, excluding the conditional variables a total of 206 variables remain (see Table 2 in the paper). The 2016 survey round includes 101 outcomes. Table 1 displays the number of outcomes by hypothesis. In total, 96 outcomes exactly match across both rounds. As a robustness analysis, we rerun Equation (1) and Equation (2) for both survey rounds at the family level restricting the analysis to the 96 variables that appear in both 2009 and 2016 survey rounds.

Table A12

Table 1. Non-conditional outcomes by Hypothesis

Hypothesis	2009	2016	Matching outcome in both rounds
Family A			
H1	7	6	5
H2	18	17	17
H3	15	7	7
Family B			
H4	15	10	6
H5	47	19	19
H6	25	4	4
H7	12	8	8
H8	15	9	9
H9	17	4	4
H10	18	9	9
H11	8	4	4
H12	9	4	4
Total	**206**	**101**	**96**

Heterogeneous Treatment Effects: We will test for heterogeneous treatment effects along the same eight community-level dimensions we used (and measured) in our earlier analysis (total households, war exposure, average schooling, distance to road, historical domestic slavery, district, ethnic fractionalization and chiefly authority). As an exploratory exercise, we will use an automated process (LASSO and BART) to identify other dimensions that are correlated with heterogeneous effects to mine the data in a principled way.

Part II: Managerial Capital

Component Overview: To evaluate a technocratic alternative to CDD's intensive social facilitation model, we overlaid a new randomized experiment across the GoBifo treatment arms. We will test whether i) a more technocratic approach to identifying project leaders with high managerial capital, and ii) the provision of training in project management fundamentals, improves community ability to active collectively and take advantage of a new development opportunity. Specifically, all communities had an opportunity to

enter a project challenge competition run by the local District Councils that awarded US$2,000 implementation grants to the twenty best project proposals. We block randomized 80 communities to a management selection treatment arm (*MS*); 78 to a management selection plus training arm (*MST*); and 80 to a control or status quo (*SQ*) mechanism that favors the village headmen.

These three treatment arms were implemented by the research team enumerators on the data collection visits to communities at the end of the focus group discussion. In all three arms, enumerators explained the project challenge opportunity and the skills needed to develop a strong proposal. They asked the group to deliberate and nominate five individuals, in addition to the village headman, who had these skills. These 6 individuals were then asked to take a management test, in private, which was scored on site by enumerators. The focus group was then reconvened and a public lottery (implemented on a tablet device) determined treatment assignment for the village. In the *status quo (SQ) arm*, the village headman was designated as the project proposal leader. His name was written on the standardized project application form and he was given a transportation voucher to redeem if/when he submitted a proposal to the relevant Local Council. In the *manager selection (MS) arm*, the enumerators announced who was the highest test performer (of the 5 non-chief nominees), and designated that person on the submission form and provided the transport voucher. The *manager selection plus training (MSTR) arm* followed the same format as *MS* but also announced that the relevant ward development committee (most local tier of elected government) would hold a one day management training as part of the project challenge competition. Enumerators provided the date and location of the training, informed the group that the travel costs of the designated project leader will be reimbursed, and encouraged the designated project leader to attend the training.

The training sessions for *MSTR* covered: i) identification of local development needs and designing projects to address them; ii) costing local materials and developing itemized budgets; and iii) time management and planning to meet deadlines. Note that measures of proposal quality capture both items covered in the training and those that were not, to evaluate the extent to which any observed training effects reflect "teaching to the test."

Hypotheses: We plan to evaluate the following hypotheses:

- There is underutilized managerial capital in villages (H-II.1)
- Leveraging underutilized managerial capital leads to greater ability to act collectively and take advantage of local development opportunities (H-II.2)
- Lack of management skills constrains the ability to take advantage of local development opportunities (H-II.3).

Econometric Specifications: Our primary tests of interest estimate:

$$P_c = \delta_0 + \delta_1 MS_c + \delta_2 TR_c + W'_c \Psi + \zeta_c \quad (3)$$

In Table 1, Panel B

where outcome *P* (i.e. proposal quality, test score of project leader) is measured for community *c*; *MS* is an indicator variable equal to one for assignment to the manager selection process (*MS* and *MSTR* arms) and zero otherwise; TR is an indicator for assignment to training (*MSTR* arm); W_c is a stratification fixed effect for geographic wards; and ζ_c the idiosyncratic error term. Hypotheses H-II.1 and H-II.2 test $\delta_1 = 0$. Hypothesis H-II.3 tests $\delta_2 = 0$.

Deviation: *Wc* is ward crossed with CDD assignment, see footnote 7

For Hypothesis H-II.1 we have only one outcome, the test score of the project proposal leader. For Hypotheses H-II.2 and H-II.3 we have four measures of proposal quality so our primary specification will

be a mean effects index. We will also report estimates for the individual scores. As a robustness check, we will exclude quality assessments that involve any input from GoBifo staff (although note all proposals were blinded during the review).

Several additional analyses will aid in interpreting these results (see [PAP Sheet 2] for details). We will:

1. Explore the extent to which the training reflects "teaching to the test." Explore where the training appears most effective. Table A8
2. Validate the management test by correlating test scores with proposal quality and explore relative predictive of power of subsection scores. Page 17
3. Validate the extent to which the distinct manager selection treatment arms translated into differences in who actually managed the project proposal process. Table 3
4. Compare the tests scores of the non-headman nominees to those of village headmen. Table 2
5. Evaluate which characteristics correlate with managerial capital test scores (i.e. age, gender, education, management experience, leadership position, etc.). Table 2
6. Test for heterogeneous response to training by management test score.
7. Test for interaction effects between participation in GoBifo and the *MS* and *TR* terms in Equation 3, noting that these tests are likely underpowered. Table 1 Panel A; Table A4

Measurement and Survey Instruments: We used several instruments to implement and evaluate this new SCA, see ["Managerial capital test", "Manager selection tally sheet enumerator A and B", "Submission survey", "Submission form", "Technical scoring", "Policy Scoring", "Expert Scoring" and data from the transcripts of the training].

Outcomes: See ["PAP, Sheet 2"]

Part III: Expert Beliefs

Component Overview: There have now been several randomized control trials of CDD projects in different countries, most of which find some positive impacts on economic outcomes and little effect on institutions. A key unanswered question is whether experts—in academia and more importantly in policy—are updating their beliefs about how effective CDD projects are. This is important in light of the large amounts of foreign aid at stake ($85 billion spent on CDD in about two decades by the World Bank alone, according to Mansuri and Rao 2012), and whether the accumulation of evidence impacts the allocation of donor funds. We surveyed students, academic and policy experts to elicit their beliefs (following DellaVigna and Pope 2016) about the long run effects of the Sierra Leone CDD project and to forecast how well communities will perform in the new project competition.[5]

We fielded this survey among several distinct groups of experts: i) policy makers working for multilateral aid agencies (including the World Bank, DfID, UNDP and IRC); ii) policy makers in Sierra Leone with knowledge of the GoBifo project; iii) economics graduate students in the US (at UC Berkeley) and the Netherlands (at Wageningen University); iv) economics undergraduate students in Sierra Leone (Fourah Bay College), v) researchers directly involved in evaluating CDD projects other development (economics) researchers; and vi) the PIs of this study. There were two versions of the survey: version 1 provided detailed information on our medium run results and version 2 asked the respondent to make predictions without any

[5] DellaVigna, S. and D. Pope, "Predicting Experimental Results: Who Knows What?" NBER Working Paper No. 22566, August 2016. See also Humphreys, M., R. Sanchez de la Sierra and P. van der Windt (2016) Social Engineering in the Tropics: A Grassroots Democratization Experiment in Congo, working paper.

information provided. For the majority of respondents, we randomized whether they completed version 1 or 2. A small subset completed both versions.

Hypotheses:

- Estimated long run treatment effects are not the same as the average prior beliefs of surveyed experts (H-III.1) Figure 4
- Average prior beliefs and forecast accuracy differ across groups of experts (H-III.2) Figure 4
- Prior beliefs about long run effects of the GoBifo project are more optimistic (e.g. predict larger positive long run effects) amongst policy makers compared to researchers (H-III.3) Figure 4
- Predictions under version 1 of the survey (that contains information on the medium run effects) are more accurate than under version 2 (H-III.4) Figure 4

Econometric Specifications: For Hypothesis H-III.1, we will evaluate whether the average prior belief across all six groups of experts are statistically distinguishable from the estimated long run treatment effects by GoBifo family and hypothesis. For H-III.2 we will test whether mean predicted effect size by family varies across groups, and assess which estimate is closest to the observed long run effects. H-III.3 tests whether the mean prior of expert groups i and ii more optimistic (predict large positive effects) than that of groups v and vi, at the family level (one sided test). Tests of H-III.4 whether prior beliefs are more accurate in version 2 compared to version 1 across all six groups. For H-III.4 we will use all the data. As a robustness check we will drop data from the subset of respondents that completed both versions of the survey.

We will run several additional descriptive analyses. These include testing whether respondents who report higher confidence in their estimates, and greater familiarity with the 2012 study, are more accurate in their predictions. For the new SCA project challenge, we will impute several estimates—regarding GoBifo treatment effects, the efficacy of training, and the impact of technocratic manager selection—and compare their mean values and accuracy across expert respondent groups.[6]

Measurement and Survey Instruments: See ["Expert Priors Survey"]

Outcomes: See ["PAP, sheet 3"].

Part IV: Impacts on Ebola

Component Overview: The recent outbreak of Ebola Virus Disease (EVD) in West Africa is the largest ever recorded. The crisis resulted in over 4000 deaths in Sierra Leone alone (about 11000 in total). The two districts where GoBifo was implemented were differentially effected, Bombali saw 1050 suspected cases and 391 deaths, while Bonthe was much less hit, with 5 suspected cases and 5 deaths. In addition to Communities suffered directly due to fear, illness and loss of life, and indirectly due to travel and trade restrictions resulting from imposed quarantines. The Ebola crisis provided a huge stress on communities at social, political and economic levels. We analyze if participation in Gobifo put communities in a better position to implement preventative measures and collaborate with local government. We report two secondary outcomes (i) we separate impacts on knowledge and collective action, and (ii) we investigate if Gobifo villages reported different Ebola case-loads.

[6] We exclude the study PIs (group vi) from this comparison. While the PIs had no access to the data, we did learn through communication with the field team that the number of submitted proposals was very high.

Hypothesis: Our main hypothesis is that "Participation in GoBifo increased knowledge, collective action and investments in preventative measures during the Ebola crisis". Table A13

Econometric Specifications: same as Equation (1) above. Our dependent variable is a mean effects index of all Ebola related outcomes. As secondary outcomes, we assess impacts in a mean effects index for knowledge and collective action outcomes separately.

We asses outcomes for the whole sample and restrict our sample to Bombali, which saw many more Ebola cases than Bonthe making the collective action outcomes more relevant.

To further interpret the hypothesis-level results, we will also estimate Equation (1) at the level of individual outcome, adjusting for FDR across outcomes. Note that this reporting of all individual outcomes is for illustrative and interpretation purposes only.

Measurement and survey instruments: see ["SES - Endline 2016", module J and K].

Outcomes: See ["PAP, sheet 4"].