

NBER WORKING PAPER SERIES

SKILL VERSUS VOICE IN LOCAL DEVELOPMENT

Katherine Casey
Rachel Glennerster
Edward Miguel
Maarten Voors

Working Paper 25022
<http://www.nber.org/papers/w25022>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2018

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.socialsciceregistry.org/trials/1784>. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research or the Department for International Development.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2018 by Katherine Casey, Rachel Glennerster, Edward Miguel, and Maarten Voors. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Skill versus Voice in Local Development

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

NBER Working Paper No. 25022

September 2018

JEL No. H41,I25,O15

ABSTRACT

Where the state is weak, traditional authorities often control the local provision of land, justice, and public goods. These authorities are criticized for ruling in an undemocratic and unaccountable fashion, and are typically quite old and poorly educated relative to younger cohorts who have benefited from recent schooling expansions. We experimentally evaluate two solutions to these problems in rural Sierra Leone: an expensive long-term intervention to make local institutions more inclusive; and a low-cost test to rapidly identify skilled technocrats and delegate project management to them. In a real-world competition for local infrastructure grants, we find that technocratic selection dominates both the status quo of chiefly control and the institutional reform intervention, leading to an average gain of one standard deviation unit in competition outcomes. The results uncover a broader failure of traditional autocratic institutions to fully exploit the human capital present in their communities. We compare these findings to the prior beliefs of experts on likely impacts, and discuss implications for competing views on the sustainability of foreign aid.

Katherine Casey
Stanford University
Graduate School of Business
655 Knight Way
Stanford, CA 94305
and NBER
kecasey@stanford.edu

Edward Miguel
Department of Economics
University of California, Berkeley
530 Evans Hall #3880
Berkeley, CA 94720
and NBER
emiguel@econ.berkeley.edu

Rachel Glennerster
Department for International Development
22 Whitehall
London, SW1A 2EG
United Kingdom
Rglennerster@gmail.com

Maarten Voors
Development Economics Group
Wageningen University
Hollandseweg 1, 6706 KN, Wageningen
The Netherlands
maarten.voors@wur.nl

A data appendix is available at
<http://www.nber.org/data-appendix/w25022>

A randomized controlled trials registry entry is available at
<https://www.socialscisceregistry.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?

This paper examines two randomized controlled trials, each designed to tackle one of these key limitations facing traditional authorities in rural Sierra Leone.

One intervention is an intensive long-run effort to make local institutions more inclusive and democratic. Such reforms are motivated by the evidence that stresses how good institutional

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>).

²Central Statistics Office (1985) for the educational attainment among those 5 years and older in 1985, and World Bank (2017) for current school enrollment rates.

design choices—those that impose checks and balances, allocate political power more equitably, and protect property rights (Engerman and Sokoloff 1997, Acemoglu, Johnson and Robinson 2001, Banerjee and Iyer 2005)—may promote long-run economic growth. By contrast, chiefs in Sierra Leone and elsewhere face few constraints on their power and rule for life (Baldwin 2016, Bulte, Richards and Voors 2018). The international aid community’s approach to reforming such autocratic institutions is to promote popular participation: giving 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 a more 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 local institutional reform (Casey 2018, pg. 145).

The other intervention we examine draws on the literature stressing 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). In particular, we study a low-cost approach that uses objective written tests to identify high skill community members, and then delegates the management of development project implementation to them. This focus on technocratic selection 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. 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.

We evaluate these two distinct approaches and compare them to each other and to the status quo of chiefly dominance, in the context of a real-world grant competition run by the district

³ Independent Evaluation Group, World Bank:
https://ieg.worldbankgroup.org/Data/reports/lp_genderincdd_01272017.pdf

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 grant competition dominates both the status quo of chiefly authority and the long-run CDD institutional reform. 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. Outcomes for the CDD communities are statistically indistinguishable from the status quo, despite 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 succeeded in generating better outcomes, and why local institutional reform largely failed.

First, we find that status quo villages do not effectively match local skills to tasks: 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, as in this case, where grant funding was on the line. This has implications beyond this specific experiment, as it suggests that traditional authorities are not adequately adapting to the large positive shock to human capital that has occurred in recent years. The skills of younger, more educated cohorts are thus considerably underutilized in the status quo approach to local development.

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.

Compared to attempts to reform local institutions, technocratic selection appears to be a promising and relatively inexpensive way to improve local project management. Yet the evidence presented here does not provide a definitive test of the broader debate regarding the primacy of human capital versus institutions, for one important reason: despite the intensity of the long-run CDD reform program, it created few detectable changes in local institutional quality or performance. Along several other measures, including the use of democratic processes in public deliberation and the local response to the 2014 Ebola public health crisis, CDD communities show no statistically significant improvement relative to the status quo control villages. CDD communities were also largely unsuccessful at bringing high skill individuals into local public

service: while we do find that chiefs in CDD villages are slightly more likely to delegate to a high skill resident, this small shift does not meaningfully affect performance in the grants competition. These largely null results mirror our earlier exploration of short- to medium-run effects (Casey, Glennerster and Miguel 2012), and further suggest that these local autocratic authorities were quite resistant to the additional years of CDD program activities that were implemented since that study. This failure reflects the perhaps unsurprising reality that it is extremely difficult to reform entrenched political hierarchies with external interventions (Ostrom 1990, 2000).

This is not to say that the overall CDD package of support, which includes both project funding and efforts to promote inclusion, is entirely 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. Given the challenges of the post-conflict operating environment, the durability of these material benefits is impressive.

Experts in public policy and academia held divergent prior beliefs over both the long-run effects of CDD and the relative efficacy of technocratic selection versus institutional reform, beliefs we elicited before analyzing the data (following DellaVigna and Pope 2018). Combining their predictions with the empirical results allows us to assess the accuracy of different types of experts over a variety of outcomes, and sheds light on competing views about the sustainability of foreign aid. Policymakers in Sierra Leone predicted that combining infrastructure grants with institutional reform would sustainably alter the local political and economic trajectory: this turned out not to be the case. However, experts from both policy and academia on average accurately predicted the durability of infrastructure investments, which is consistent with a more modest view of sustainability. Here CDD compares favorably to other foreign aid approaches that are associated with rapidly deteriorating public infrastructure (see Miguel and Gugerty 2005). All expert opinion was inaccurate in some dimensions: every expert underestimated the extent of community entry into the project grants competition, a result that underscores the intense demand for public goods and the pervasive barriers to accessing financial capital in these communities. It further suggests that the CDD villages remain a long way from sustainability as understood by no longer needing financial subsidies (for related arguments in health, see Kremer and Miguel 2007, Dupas 2014).

The rest of the paper is structured as follows. Section II describes the experimental design and details of the two interventions. Section III presents the main empirical findings, and Sections IV and V explore mechanisms driving these results. Section VI considers project sustainability

and the accuracy of expert priors, and the final section concludes with discussion of policy implications and external validity.

II. Experimental Design and Interventions

This research was designed around a real-world economic development opportunity. In 2016, the local elected governments in our Sierra Leone study districts 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 (20) proposals. 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.

We compare the effectiveness of two distinct interventions in allowing communities to avail themselves of this new opportunity, and benchmark both against the status quo 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 grant 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 project challenge 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 project challenge 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 two things: first, identifying individuals with the appropriate skills; and second, encouraging community leaders to delegate project management authority to them.

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 each study community, focused on publicizing the district government grants competition. The team supervisor explained the size of the grants, how the competition worked, and encouraged communities to enter. Then he (or she) carefully 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), the performance of public agencies in Nigeria (Rasul and Rogger 2016), and the 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 named 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 different public “nudges.” 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 nudge towards delegation to the chief is the control condition. The supervisor then made a public 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 these announced nudges were public, there is nothing binding about the project’s encouragement to delegate to the technocrat or the chief. There are several reasons to believe that these nudges would have little effect on the nature of project delegation. First, the communities were informed that the project grant 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 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.

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 receive from the research team. Finally, if the traditional chiefs see these high human capital managers as a political threat, they may choose to sideline them from the process or sabotage their efforts. Any combination of these four factors would work against finding a treatment effect of the technocratic selection nudge on performance in the grants competition.

There is one additional finding from a well-known study of Indian firms that we elected to test in this setting, namely, whether managerial practices can be effectively taught (Bloom et al. 2013). To do so, we subsidized the cost of attending a one day training course in basic project management skills (e.g., budgeting, planning, and grant writing) for a randomly chosen half of the selected technocrats in treatment villages. (No training was offered in the status quo selection communities, see Figure 1.) These trainings 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. In these randomly chosen training 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 treatment 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 has two main elements: block grants provided to communities to fund public infrastructure; and intensive social facilitation to promote more inclusive and accountable local governance. Program activities typically began by establishing a village development committee (VDC), mandated to include representatives of marginalized groups, which is trained and encouraged to make the selection, planning and implementation of community projects in a highly participatory 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. The idea is that this investment in organizational capacity could permanently lower the fixed cost of collective action—which can be applied to future community decisions and development activities—and thus ideally places communities on a stronger development trajectory that outlasts the direct financing stage. We test this claim by seeing whether the chiefs and other community leaders that have 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 project challenge 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 also 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 costs are comparable for the second, less active phase (2010-present), at nearly \$3 million, given the continuation of project staffing, transport and overhead. The relatively high cost of the institution building CDD component serves as further motivation for the technocratic selection intervention, which is far less expensive and 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, social capital, and inclusive governance.

These null results on institutional outcomes 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 medium run data collection (in 2009), in order to assess long-run institutional change. Note further that the 2016 survey round incorporates the additional “dose” of CDD funding and programming that began implementation in 2010.

Given the divergent perspectives and interpretations of the medium-run results, we thought it important to systematically document the prior beliefs of experts in both the policy realm and in academia before analyzing the long-run follow-up data (see Appendix B for details). To do so, we surveyed 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; and faculty in both economics and political science who have been directly involved in evaluating CDD projects or working in related areas of development economics (including ourselves, the co-authors of this article).

We asked these experts to predict long-run treatment effects for the same set of institutional indicators we used previously. They were also surveyed about their prior beliefs regarding the persistence of infrastructure and economic activity gains, as well as expectations regarding entry in the district government grants competition. We randomly varied whether the survey prompted the expert with the medium run results (from Casey et al. 2012).

In total, we collected priors from 48 experts and Figure 2 plots their predictions along key dimensions (in panels A through C), where the circles denote individual expert predictions and the whisker plot portrays the distribution.⁴

We start by focusing on beliefs about the long-run impacts of CDD on institutional

⁴ We also surveyed 78 economics students in both Sierra Leone and the OECD, whose priors closely track those of the experts in their respective regions, see Appendix B.

outcomes, and below discuss these expert priors and their relationship to estimated long-run treatment effects in further detail. It is clear that there is wide divergence in the beliefs of experts both across the categories and within each group. Policymakers in Sierra Leone predicted average long-run effects in the range of 0.25 standard deviation units, which is significantly larger than the 0.06 effect predicted by policymakers in the OECD. While faculty appear more pessimistic (mean expectation is a 0.03 units effect), their predictions are not statistically distinct from the OECD policy experts. Pooled together, the experts predicted a long-run effect that significantly exceeds what Casey et al. (2012) estimated in the medium run (0.095 predicted by experts, compared to 0.028 standard deviation units), and remains statistically different than zero even when limited to the random subgroup of experts that were primed with information on the medium-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 question.

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 project challenge competition (arm 2) in comparison to that of autocratic chiefs in the status quo (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 institutional reform (in arms 5 and 6).

III. Main Empirical Results

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

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

where outcome P (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 institutional reform, respectively, to the status quo 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 G 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 and grant award decisions made by the district governments themselves. As we have the complete ranking of all submissions, we test for impact on winning a grant for the actual budgetary cut off, as well as for lower thresholds that would have obtained if the government had had more funds to allocate. 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.

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 deviations (standard error 0.164) for technocratic selection, indicating that communities nudged to choose a high skill manager submitted proposals of substantially higher quality than the chiefs in status quo villages (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.⁵ 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,

⁵ 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.

as compared to traditional chiefly dominance.

The five analogous treatment effect estimates of CDD are much smaller in magnitude and none are statistically distinguishable from zero at conventional levels, indicating that participation in the multi-year institutional reform intervention did not substantially alter community ability to access a new funding opportunity that shares many features with what they were trained in during the GoBifo project. Estimates in the third row provide no evidence for significant interaction effects between technocratic selection and institutional reform. (For alternative specifications, see Appendix Table A4 for the fully interacted model and Table A5 for a simple two-way comparison between CDD and technical selection with no interaction terms.)

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:

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

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).⁶

Results are presented in Table 1, Panel B. The estimated treatment effect for technocratic selection alone is a 0.315 standard deviation unit improvement in proposal quality (standard error 0.138), as compared to project management under the status quo 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).⁷ 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 unit 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

⁶ 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.

⁷ In total, two people assigned to the training did not show up and four people not assigned were trained.

the technocrats' proposals were of higher quality, this did not significantly affect whether or not communities won an implementation grant in this regression specification: estimates in column 5 (of Table 1, Panel B) are positive but not statistically distinct from zero (0.067 with standard error 0.044). This partially reflects the high take up rates – so many communities entered the competition that the baseline rate of winning a grant was only 8.8 percent – and accompanying reduction in statistical power when parsing the effects of training from selection.

As statistical power is restricted by the low number of grants awarded, and the number of grants is somewhat arbitrary, an informative thought exercise is whether there would have been treatment effects on winning a grant had the government budget for the exercise been larger. Figure 3 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, which is very competitive. If we relax this, there are other lower thresholds for which we would detect a treatment effect on winning a grant (see F -tests in Appendix Table A6).

Figure 4 summarizes some of these main results by plotting the mean proposal quality score for several different types of villages: those in the status quo (treatment arm 1 from Figure 1), villages that experienced long-term institutional reform under CDD (arm 4), villages nudged to delegate to a high skill manager but without management training and no CDD experience (arm 2), and CDD control villages nudged to delegate to a high skill manager who also received management training (arm 3). Average performance is not affected significantly by several years of CDD exposure (p -value = 0.61). Technocratic selection villages outperform status quo villages (by 0.35 standard deviation units, p -value = 0.10), and they benefit further from subsidized management training, with a combined effect of 0.50 standard deviation units (p -value = 0.02).

These differences are particularly important in light of the substantial cost differences across interventions. 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 the cost of institutional reform is a full order of magnitude greater than the technocratic interventions, and took years to implement, in contrast to a few days.

This cost comparison warrants one important observation. The technocratic selection approach 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.

Section IV: Why Technocratic Selection Worked

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 successfully identified high skill individuals. Comparing technocratic selection to the chiefly status quo, 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.⁸

Second, the written test scores are informative of performance in the project challenge grants competition. There is a positive correlation between the score of the selected project manager (whether chief or top scorer) and outcomes in the district government grants 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).

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

⁸ Note that we estimate a null result on whether management training further enhanced the technocrats' scores (-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.

(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 male.

The third step in the causal chain requires that the public nudges to choose the high skill individual increase the likelihood that they are in fact put in charge of managing the community's entry into the grant 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. We asked for detailed information about who was involved in selecting which type of project to apply for, developing the budget, and setting the implementation timeline. 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). Note the presence of substantial two-sided non-compliance with the project manager nudge: in 20 percent of status quo communities, 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 and the chiefly authority in charge in the control villages, there was no meaningful constraint on communities reverting to the status quo 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.⁹

⁹ 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.

Yet 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 status quo 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 status quo arm. 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.

Turning to the next intervention, why was management training 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 A7 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, a field team 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 these 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 A8). 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. That said, there are hints of support for the hypothesis that chiefs are better able to get things done, as the rates of functionality, quality and contributions are all higher for status quo selection. However, this is based on the four 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 42 percent.

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 jobs. 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 2015, Deserranno 2017). 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 Institutional Reform Largely Failed

The most immediate explanation for why institutional reform did not improve community

performance in the project challenge competition is that we find scant evidence that the intensive CDD experience changed how communities are actually governed. We first trace this through each step in the grants competition and then apply it to a panel of broader institutional performance 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 project challenge. 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 mutli-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).¹⁰ This further suggests that the learning-by-doing in implementing public infrastructure 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.

¹⁰ This is for the 236 managers 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.

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 sample of technocratic selection arms.

There is also no evidence from textual analysis that the submitted 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 A9). 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 small reallocation of project work towards high skill community members was not sufficiently large to meaningfully affect performance in the grants competition.

This finding that even an intensive CDD project leads to only small improvements in governance resonates with other institutional measures we collected. During the 2016 field visits, enumerators collected data on as many of the same indicators of institutional quality and performance that we had measured in our 2009 study. Table 4, Panel A presents estimates of long-run CDD impacts on these measures. Combining all 95 individual outcomes 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 only 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 A10). 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 “zombie” 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 A11 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 (s.e. 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.

Section VI: Expert Expectations about CDD Impacts and Sustainability

Another way to gauge the magnitude of the long-run institutional impacts due to CDD is to compare them to what experts predicted. Returning to Figure 2, it is apparent that policy experts in Sierra Leone greatly overestimated the potential for long-run institutional change, while OECD policymakers were roughly on target on average (Appendix Figure A1 presents confidence bands). While we cannot reject that economics and political science faculty were correct on average, they were more pessimistic, with a substantial number of them (11 out of 23) predicting precisely zero long-run effects, which falls outside the 95% confidence interval of the observed point estimate.

The 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.

The divergence in expert views about likely long-run CDD impacts raises questions about how we think about the sustainability of foreign aid programs more broadly. An ambitious conception of sustainability is a difference in both “level” and “slope.” CDD advocates claim that

combining infrastructure grants with institutional reform provides an economic boost in the short run while also permanently reducing the fixed costs of collective action, thus facilitating future public investments (and presumably growth), see Dongier et al. (2002) for a discussion. The lack of any detectable effect on communities' ability to access the grants in the new project challenge competition runs counter to this view.

A more modest conception of sustainability involves a difference in economic levels that persists over time, outlasting the stream of direct financial support. 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 longer than other types of aid. 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. 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 our 2009 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 F.

Table 4, Panel B presents results. All 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 (in column 3). For the infrastructure family overall, 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 (standard error 0.041) and that for 2009 is 0.352 (s.e. 0.035), see Appendix Table A10.

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.¹¹ 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.¹²

While the observed rate of decay aligns closely with what the sample of experts predicted on average, this masks a great deal of heterogeneity in their expectations. Pooled together, the experts predicted a treatment effect for this family overall 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 2, 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 project grant competition. We asked experts to make predictions about community entry into the grants competition for each of the six experimental cells (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

¹¹ Regarding heterogeneity in treatment effects, we find evidence for larger impacts on public goods in smaller population communities in the long-run data, consistent with Anderson and Magruder’s (2017) reanalysis of our 2009 data. This heterogeneity could reflect greater challenges with collective action in larger groups (Olson 1965) or effectively smaller per capita grants in larger communities (GoBifo financial grants were uniform in size for all communities). See Appendix E for discussion.

¹² The lack of household survey data prevents us from directly measuring effects on wealth, consumption or income.

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 2, Panel C).

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 disagreement across expert type for institutional change, wide dispersion for the durability of infrastructure, and systematic underestimation for entry into the grants competition. 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.¹³

More broadly, the near universal entry of communities into the grants competition reflects the scarcity of financial capital and dearth of public goods in these rural Sierra Leone communities. It suggests that, even after extensive CDD programming, poor communities remain a great distance away from independence from subsidies for public goods, an argument that has already been made strongly for low-income countries in public health (Kremer and Miguel 2007, Dupas 2014).

Section VII: 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 build inclusive local governance institutions yields little in the way of impacts on communities’ ability to compete in the external grant 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 may be a more viable and affordable strategy than attempts to affect institutional transformation in Sierra Leone.

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

¹³ See Della Vigna and Pope (forthcoming), Humphreys, Sanchez de la Sierra and van der Windt (2015) and Vivaldi and Coville (2017).

had completed primary education and only 4% had completed secondary,¹⁴ which would have greatly limited the scope for recruiting high skill local 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 (see Wong 2012, King and Samii 2014, White et al. 2017, and Casey 2018 for cross-country reviews), as well as on other approaches to harness local skill even when institutional transformation is not practical.

References

- 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, Daron, 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.

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

- Anderson ML, and Magruder J. 2017. Split-Sample Strategies for Avoiding False Discoveries. Working Paper.
- Ashraf, N., O. Bandiera, and S. 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.
- Banerjee, A. and Iyer L. 2005. History, Institutions, and Economic Performance: The Legacy of Colonial Land Tenure Systems in India. *American Economic Review* 95(4):1190-1213.
- Barro, R. 1991. "Economic Growth in a Cross Section of Countries," *Quarterly Journal of Economics*, 106(2): 407-443.
- 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, V4.
- 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.
- Dupas, P. 2014. “Getting Essential Health Products to Their End Users: Subsidize, but How Much?” *Science*, 345(6202): 1279-1281.
- Engerman, SL, and KL Sokolof. 1997. Factor Endowments, Institutions, and Differential Paths of Growth Among New World Economies: A View from Economic Historians of the United States. In *How Latin America Fell Behind*. Stanford: Stanford University Press.
- Finan, F., R. Pande and B. Olken. 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.
- Humphreys M, Sanchez de la Sierra R, van der Windt P. 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.

- Miguel, E., Gugerty MK. 2005. "Ethnic Diversity, Social Sanctions, and Public Goods in Kenya," *Journal of Public Economics* 89:2325-68.
- Olson M. 1965. *The Logic of Collective Action: Public Goods and the Theory of Groups*. Cambridge: Harvard University Press.
- Ostrom E. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press.
- Ostrom E. 2000. Collective Action and the Evolution of Social Norms. *The Journal of Economic Perspectives* 14(3):137-58.
- Rasul, I., and D. Rogger. 2016. "Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service," *Economic Journal*, 128: 413–446
- 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 29.
- 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, Michael. 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

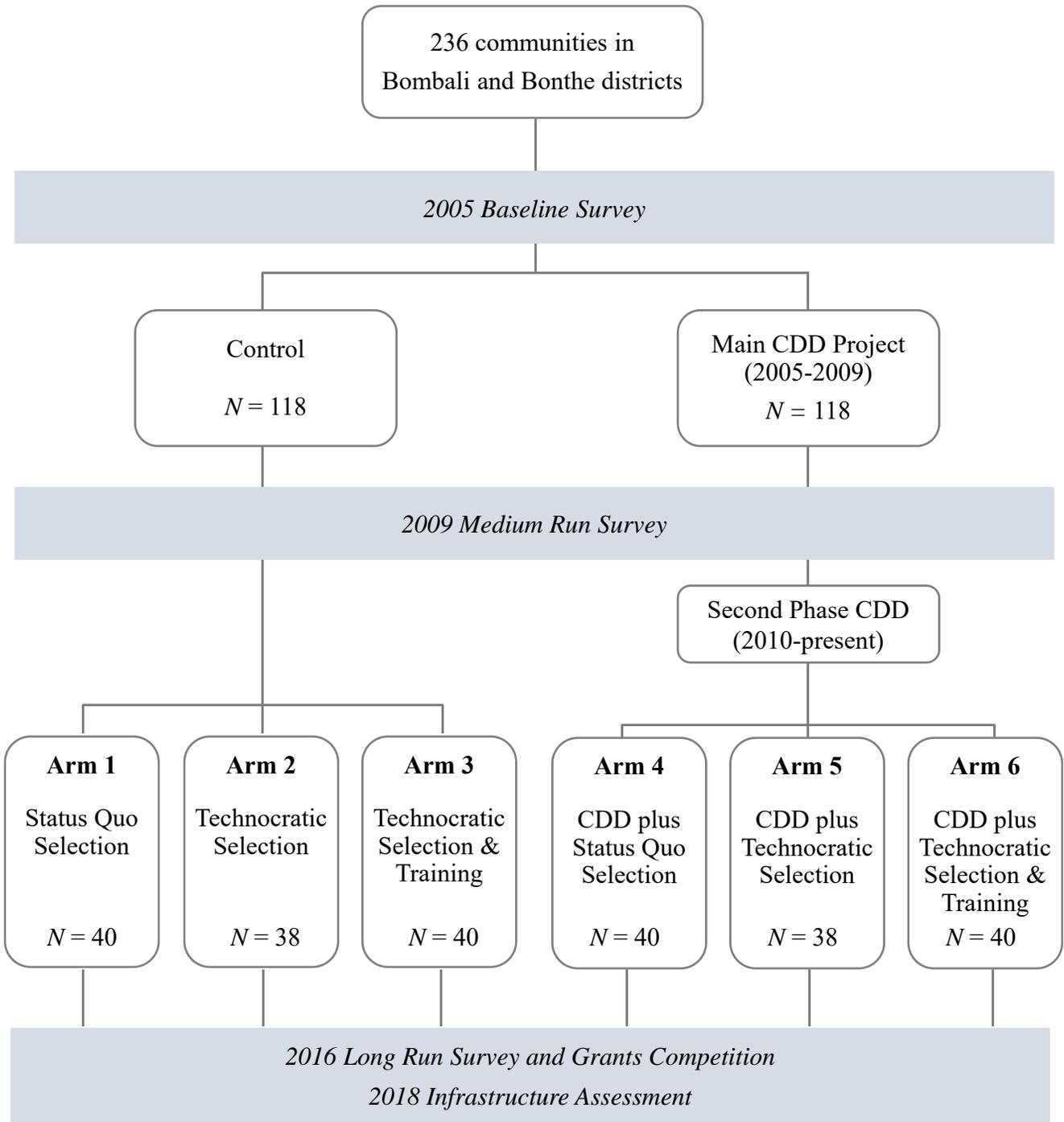
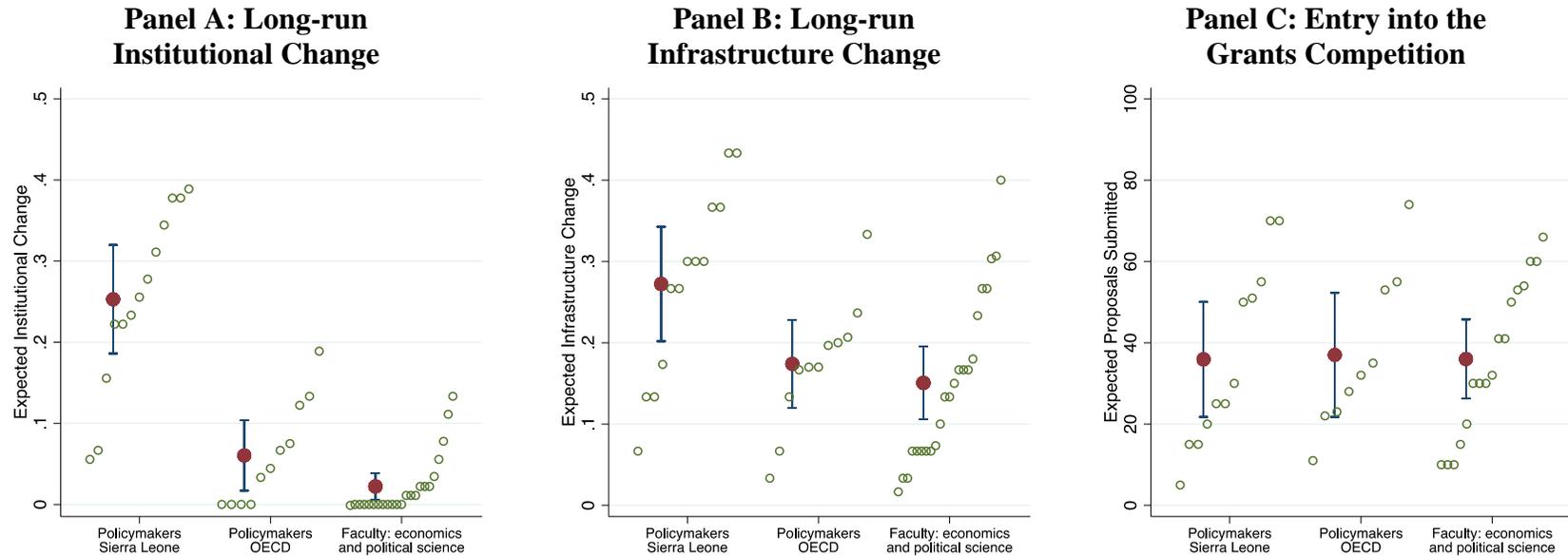
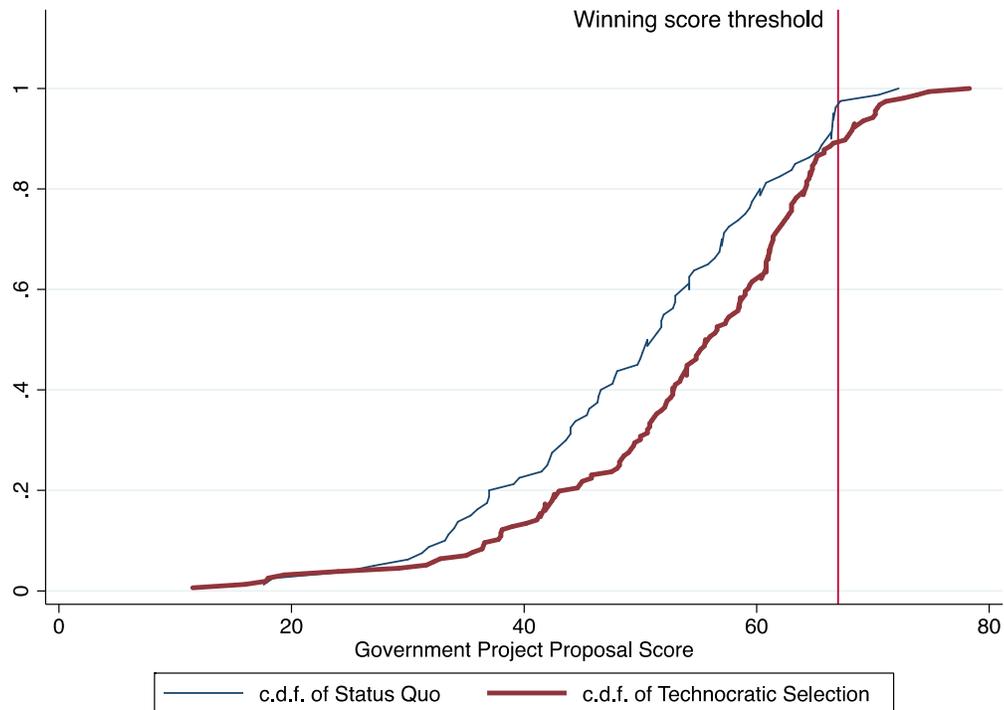


Figure 2: Expert Predictions of Long-run CDD Effects and Grants Competition



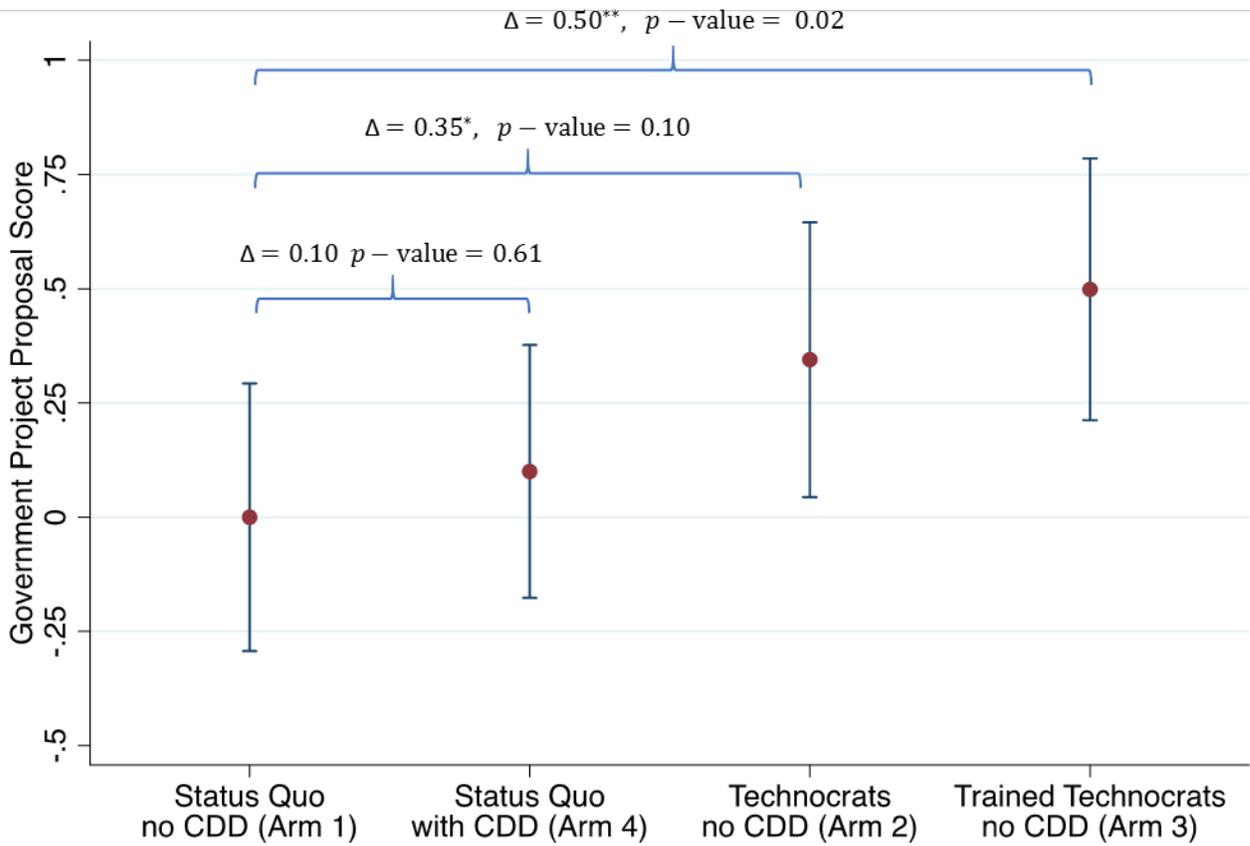
Notes: This figure presents predictions from 48 experts collected during December 2016 and July 2017 before any data analysis. Panels A and B present expectations for CDD treatment effects measured in standard deviation units. Panel C presents expectations about the percent of communities in the base case (no CDD, status quo chiefly control, or arm 1 of Figure 1) that would enter the grants competition (i.e. submit a proposal). 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.

Figure 3: 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 status quo selection (arms 1 and 4 in Figure 1) and to technocratic selection (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 treatment arm mean for the four non-submitting communities ($N = 236$). A Kolmogorov-Smirnov test rejects equivalence of the two distributions at $p\text{-value} = 0.03$.

Figure 4: Proposal Performance across Manager Selection Arms



Notes: This figure presents the mean proposal score index for different types of managers selected in each of the four experimental arms indicated in Figure 1. Missing values for communities that did not submit a proposal are imputed at the treatment arm mean.

Table 1: Treatment Effects on Performance in the Grants Competition

	Proposal Score (index) (1)	Technical Score (2)	Expert Score (3)	Gov't Score (4)	Won a Grant (5)
Panel A: Technocratic Selection versus CDD Institutional Reform					
Technocratic Selection	0.397** (0.164)	0.526*** (0.193)	0.377** (0.169)	0.289 (0.177)	0.101** (0.049)
CDD	0.061 (0.181)	-0.015 (0.206)	0.063 (0.192)	0.136 (0.190)	0.048 (0.047)
Technocratic Selection * CDD	0.094 (0.222)	0.017 (0.255)	0.218 (0.232)	0.047 (0.238)	-0.098 (0.067)
Observations	236	236	236	236	236
<i>F</i> -statistic (on TS and TS*CDD)	8.00	8.65	9.01	3.44	2.11
<i>p</i> -value	<0.001	<0.001	<0.001	0.034	0.124
Panel B: Technocratic Selection and Managerial Training					
Technocratic Selection	0.315** (0.138)	0.435*** (0.156)	0.298** (0.140)	0.214 (0.152)	0.067 (0.044)
Training	0.339** (0.133)	0.280* (0.157)	0.446*** (0.130)	0.292* (0.155)	-0.026 (0.048)
<i>F</i> -statistic (on TS and TR)	12.59	11.61	16.09	5.86	1.28
<i>p</i> -value	<0.001	<0.001	<0.001	0.003	0.281
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) the 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	<i>p</i> -value on difference	<i>N</i>
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)	<i>p</i> -value on difference	<i>N</i>
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	Status Quo Selection (arms 1, 4)	Technocratic Selection (arms 2, 3, 5, 6)	<i>p</i> -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)	<i>p</i> -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)	<i>p</i> -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 proportion 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 status quo 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 <i>p</i> -value (2)	FDR adjusted <i>q</i> -value (3)	Treatment effect 2009 (4)	Change over time (1) - (4)
Panel A: Institutions Family					
All outcomes (61 unique outcomes)	0.066*** (0.025)	<0.01	0.005	0.028 (0.020)	0.037 (0.028)
Collective action	0.098 (0.050)	0.049	0.235	0.012 (0.037)	0.086 (0.061)
Inclusion	0.033 (0.036)	0.350	0.539	0.002 (0.032)	0.031 (0.044)
Local authority	-0.035 (0.068)	0.604	0.632	0.056 (0.037)	-0.088 (0.070)
Trust	0.107 (0.057)	0.063	0.235	0.042 (0.046)	0.064 (0.081)
Groups and networks	0.149 (0.071)	0.037	0.235	0.028 (0.037)	0.121 (0.074)
Access to information	-0.036 (0.067)	0.590	0.632	0.038 (0.037)	-0.075 (0.072)
Participation in governance	0.079 (0.060)	0.190	0.348	0.090*** (0.045)	-0.011 (0.065)
Crime and conflict	-0.002 (0.063)	0.971	0.759	0.01 (0.043)	-0.012 (0.074)
Political and social attitudes	0.154 (0.124)	0.215	0.348	0.041 (0.043)	0.113 (0.126)
Panel B: Infrastructure Family					
All outcomes (30 unique outcomes)	0.204*** (0.040)	<0.01	0.001	0.298*** (0.031)	-0.094*** (0.036)
Project implementation	0.253*** (0.068)	<0.01	0.001	0.703*** (0.055)	-0.450*** (0.081)
Local public goods	0.228*** (0.046)	<0.01	0.001	0.204*** (0.039)	0.024 (0.041)
Economic welfare	0.240*** (0.056)	<0.01	0.001	0.376*** (0.047)	-0.136** (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).