Can Communities Take Charge?

A Randomized Controlled Trial on Sustaining Schools in Afghanistan*

Dana Burde, Joel Middleton, Roxanne Rahnama, and Cyrus Samii†

February 15, 2023

*This project is generously supported by the US Agency for International Development and the Canadian International Development Agency and Global Affairs Canada. The project was reviewed and approved by the New York University institutional review board as study IRB-FY2016-733. The pre-analysis plan was registered with EGAP as study 20180817AA prior to viewing any outcome data. We are grateful for feedback from participants at APSA 2020, Boston University, University of Chicago, EGAP, Hertie School, ITAM, the Petralia Workshop, and NYU Development Research Institute.

†Burde is Associate Professor at New York University (dana.burde@nyu.edu). Middleton is Assistant Professor at the University of California, Berkeley (joel.middleton@berkeley.edu). Rahnama is PhD Student at New York University (rr3587@nyu.edu). Samii (corresponding author) is Associate Professor at New York University (cds2083@nyu.edu).

Can Communities Take Charge?

A Randomized Controlled Trial on Sustaining Schools in Afghanistan

Abstract

After wars, international nongovernmental organizations (INGOs) often service basic needs like health and education. Ideally, such service provision would eventually be handed over to an accountable government. But cases like Afghanistan show that handover can fail. Should the ideal of handover be abandoned, or are there strategies to make it realistic? To address this question, we test a community-based model for administering primary education in Afghanistan. We use a randomized equivalence trial that estimates effects relative to what INGOs achieve. The model succeeds in maintaining children's learning and household satisfaction. It does not increase gender or ethnic bias nor does it result in resource misappropriation. However, survey results from a one-year follow-up with community leaders suggests that long-term viability of the model depends on higher level institutional supports being in place. Communities can take charge in administering basic services after INGOs withdraw, but only when higher level institutions are stable.

1 Introduction

In the aftermath of war, international non-governmental organizations (INGOs) often service basic needs. Since the 1990s, INGOs have proliferated in global presence, directing increasing shares of foreign aid (Banks, Hulme and Edwards, 2015). Resources disbursed for INGO projects was estimated to be around $0.9 billion in 1970, and this number rose to $6.3 billion by 1993 and continued to grow subsequently (Werker and Ahmed, 2008). Nonetheless, there are longstanding concerns over the sustainability of large-scale service provision by INGOs (Edwards and Hulme, 1996), which is often motivated as a "stopgap until the government can provide similar services in the long-run" (Deserranno, Nansamba and Qian, 2020).

A question for policy and for the statebuilding literature concerns whether and how to hand over service provision from INGOs to governments. A normative reason for such handover is to ensure that service providers are directly accountable to those that they are serving (Nussbaum, 2019; Tanasoca and Dryzek, 2021; Rubenstein, 2015). Scholars voice concerns over the possible "rewriting of the social contract between government and its citizens as a result of INGO substitution for the state" (Edwards and Hulme, 1996, p. 967; see also Campbell, DiGiuseppe and Murdie, 2019). But practically, if governments cannot meet the performance standards of INGOs, handover to government could reduce citizens' material welfare. Current literature shows handover to be difficult, with barriers to mainstreaming interventions introduced by INGOs (Bold et al., 2018; Banerjee et al., 2017; Raffler, Posner and Parkerson, 2019). If citizens are used to INGOs high performance standards, to have the government step in and fail could be perilous. Balancing accountability with efficacy is a general governance dilemma (Gadenne and Singhal, 2014; Alesina and Tabellini, 2007), but it is particularly acute in war-affected contexts in which vulnerability is widespread and government legitimacy is questioned (Bandiera et al., 2019; Casey et al., 2018).

We use a 2014-19 field experiment in Afghanistan on education provision to study a strategy for managing this dilemma. This is a high-stakes policy area, given the priority that Afghans as well as citizens around the world place on access to education (Rose and

Greeley, 2006). In conflict-affected countries, INGO-run schools are demonstrably effective in improving access and learning in hard-to-reach areas (De Regt, Majumdar and Singh, 2013). In Afghanistan, these schools have produced substantial learning and attendance gains, particularly for girls (Burde and Linden, 2013), and increased trust in government and INGOs (Burde et al, in progress). However, the Afghanistan case illustrates the challenges of handover to the domestic government. A 2014 survey reported in Burde, Middleton and Samii (2016) found that three years after INGOs withdrew and transferred responsibility to the Ministry of Education in Afghanistan in 209 villages, in 77 percent of them the schools closed, no nearby-substitute was offered, and attendance reverted dramatically, especially for girls. This failure could not be attributed to lack of resources, with billions in unspent reconstruction funds leading policymakers in Afghanistan's largest external donor, the US, to question whether allocations to reconstruction were excessive (Congressional Research Service, 2020). Nor was this a matter of citizens demanding reallocation to things other than education, given very strong demand for primary education in these communities (Burde, Middleton and Samii, 2016).

A potential solution to this challenge rests in decentralized, community-based governance. We test of a strategy that makes use of a decentralized institutional infrastructure created through a previous nationwide community-driven development program in Afghanistan, the National Solidarity Program (Beath, Christia and Enikolopov, 2015). Through a 2014-19 collaboration with Afghan government agencies and two INGOs, we study a community-based "sustainability model" that has INGOs hand over managerial responsibility for primary schools to community development councils. The strategy builds on community-based school management strategies in developing countries (Ganimian, 2016; Patrinos, Barrera-Osorio and Fasih, 2009), including in post-conflict contexts such as El Salvador (Jimenez and Sawada, 1999), as well as community-level management of public goods in fragile states (Berman et al., 2017). We use a randomized "non-inferiority equivalence trial" to test outcomes against a cost-adjusted standard set by INGOs. The trial's design responded to the

ethics of the situation. We test whether community councils can manage schools in ways that reduce costs and sustain attendance and learning outcomes at least as well as, or not substantially worse than, INGOs.

The sustainability model maintains childrens' access to school and their learning (measured by math and reading tests) relative to the cost-adjusted INGO benchmark. Households were similarly satisfied with education quality in the community-managed schools as compared to INGO-run schools. The community-managed schools were no more biased on the basis of gender or ethnicity. Giving authority to community councils did not introduce consequential agency problems or resource misappropriation, contrary to conventional concerns in the decentralization literature. In fact, teachers were paid *more* regularly under community management. The model worked with both customary leaders and other elected council members, with no apparent incompatibility between these two bases of local authority (Jochem, Murtazashvili and Murtazashvili, 2016). In a follow-up survey a year after the intervention, community leaders expressed pessimism about the model's long-term viability. This suggests that for the model to be viable, leaders must be assured of stable higher-level support. Moreover, given that administrative effort displaces other activities, community leaders should be compensated for their effort.

The period of this study (2014-19) overlapped with the two terms of the Ashraf Ghani administration, and the study sites were outside the areas under Taliban control prior to their 2021 takeover. Nonetheless, it is reasonable to expect that the sustainability model would be robust to changes in national leadership. Unlike administrative networks tied into the former central government, the community councils have largely remained intact and, amazingly, have continued to serve as network for the administration of internationally-sponsored humanitarian aid *even under* Taliban rule.[1] This echoes past indication of the resilience of community institutions despite shifts in centralized authority (Murtazashvili,

[1]This assessment is based on discussions with with the Community Directed Development Organization (CDDO), which is an Afghan-centered NGO facilitating the delivery of aid in Afghanistan via the community council network, as well as with researchers and INGOs who have worked through this network.

2016).

We contribute to the literature on decentralized service provision and statebuilding. We offer a fresh take on "community-directed development" (CDD) (Casey et al., 2018; King and Samii, 2014; Mansuri and Rao, 2004; Christia, 2019). We show that CDD institutions offer a durable institutional architecture for administering services in challenging settings. Our results shed light on circumstances under which local leaders are induced to act honestly and play a constructive role (Lund and Saito-Jensen, 2013; Alatas et al., 2019; Fritzen, 2007; Dasgupta and Beard, 2007; Mansuri and Rao, 2004). Our results offer a promising approach to government scale-up of INGO-introduced programs, serving as a counterweight to more pessimistic findings in other recent studies (Bold et al., 2018; Banerjee et al., 2017; Raffler, Posner and Parkerson, 2019). This study also illustrates a decentralization-based approach to transitioning from aid to state-building (Blair and Winters, 2020).

The remainder of this article proceeds as follows: we relate our study to theories of decentralization in Section 2. We present the context in Afghanistan in Section 3, the intervention in Section 4, and the research design and ethical considerations in Sections 5, 6, and 7. We present short-term results in Sections 8, 8.4 and longer-term results in Section 9. We conclude and discuss implications in Section 10.

2 Theoretical Considerations

Both INGOs and governments can centralize administration of basic services. Resources and personnel recruitment are controlled centrally, and agents are held accountable "from above" by senior authorities. In the case of education, teachers form a front line between this centralized, top-down accountabilty chain and would-be beneficiaries (the community). Agency problems can arise between any link in a top-down accountability chain, from the leaders in the line ministry all the way to the teachers (Duflo, Dupas and Kremer, 2015). To minimize such risks, governments or INGOs often minimize the number of links in the top-down

chain. The problem is that this creates new costs: service administration requires regular local problem solving. Central-agency administrators need to assess and service changing local needs. For servicing remote communities, like in rural Afghanistan, centralization can be costly and inefficient when it comes to assessing and servicing local needs. Administrators need to travel out to local communities. They face constraints on their abilities to familiarize themselves with local concerns. Moreover, strengthening centralized capacity risks reproducing historical dynamics of dependency and expropriation that have fed conflicts in places like Afghanistan (Murtazashvili and Murtazashvili, 2020).

Given these challenges, we focus here on the potential for a *community-based* decentralization strategy for service administration that delegates authority to local community councils. The political economy literature offers reasons for both optimism and caution with such an approach. Localizing administration can improve efficiency and responsiveness to local interests (Oates, 1972; Galasso and Ravallion, 2005; Gadenne and Singhal, 2014). Numerous papers provide direct evidence for this prediction in relation to education, including pro-poor investments (Faguet, 2004; Galasso and Ravallion, 2005), enrollment and learning (Galiani, Gertler and Schargrodsky, 2008; Andrabi et al., 2008), teacher performance and regularity of teacher compensation (Bold et al., 2018; Duflo, Dupas and Kremer, 2015), and children's health (Björkman and Svensson, 2009). Such positive effects appear to be driven by both improving access to information on local preference and subjecting resource allocation processes to "bottom-up" accountability pressure through inclusive decision-making (Bardhan, 2002; Gadenne and Singhal, 2014; Beath, Christia and Enikolopov, 2017; Fox, 2015).

Localizing administration could worsen public service delivery by creating new agency problems (Bardhan and Mookherjee, 2006; Gadenne and Singhal, 2014). Decentralized administration grants local authorities discretion in personnel recruitment and resource use. If bottom-up accountability is weak, this creates risks of elite capture, resource misappropriation, or theft (Platteau and Gaspart, 2003; Reinikka and Svensson, 2004). Local leaders may

recruit personnel based on patronage rather than ability to service needs (Altschuler, 2013; Wilson, 1961; Olken, 2007). These risks can be high in developing countries with weak monitoring mechanisms (Bardhan, 2002), and in conflict-affected settings accountability may be even weaker (Christia, 2019). Local leaders may also target goods in discriminatory ways, on the basis of co-ethnicity or prioritizing males (Bardhan and Mookherjee, 2006; Chattopadhyay and Duflo, 2004). In Afghanistan, Beath, Christia and Enikolopov (2018) discuss how local leaders may react opportunistically to ambiguities in institutional accountability, in line with theories on parallel institutions and corruption (Shleifer and Vishny, 1993).

There are two ways through which the particular decentralization strategy studied here may mitigate such agency problems. First, the strategy studied here works through an institutional network created through a "community directed development" (CDD) initiative, as described below (Mansuri and Rao, 2012; Casey et al., 2018). Vulnerability to capture depends on levels of citizen mobilization and awareness (Bardhan, 2002; Platteau and Gaspart, 2003; Ensminger, 2017). Community directed development initiatives have evolved in ways to promote bottom-up accountability to citizens (Fox, 2015). Beath, Christia and Enikolopov (2018) find in Afghanistan that the CDD councils that form the basis of our strategy are accountable: embezzlement was lower in food aid distribution by elected village development councils as compared to customary institutions. Households' interest in education is widely shared by community members—indeed, Burde et al. (2019) find that the vast majority of households in communities from all major ethnicities in Afghanistan want both their boys and girls to be educated, at least up to high school. This gives leaders extrinsic motivation to improve education for the village children (Beath, Christia and Enikolopov, 2017).

Agency problems may be mitigated if leaders have an *intrinsic* interest in the service provided. For Bardhan and Mookherjee (2005), the potential for local elite capture under decentralization depends on leaders' preferences. Galiani, Gertler and Schargrodsky (2008) claim that "if local elites do not use the local public schools, they would likely prefer that

public resources were allocated elsewhere" (p. 2107). But the reverse should also hold: if local elites themselves have an interest in having their own children or their community's children educated, then embezzling funds dedicated to education would not be in their interest. However, such reliance on leaders' intrinsic interests would mean that the benefits of decentralization are contingent on leaders' interests, rather than on the institution.

Based on this analysis, a decentralization strategy that makes use of community-directed development institutions may have promise in a setting in which top-down government strategies have previously failed. However, in studying the effectiveness of the strategy, we must be attentive to potential agency problems, misappropriation, bias, or dependence on elites' particular interests. These insights inform the set of outcomes that we choose to assess in our study.

3 Institutional Context

Our study takes place in an extremely challenging institutional context. After the fall of the Taliban in 2001, the Afghanistan government faced the challenge of rebuilding after decades of war, including the need to extend services to a highly dispersed and predominantly rural population (Rashid, 2008; Suhrke, 2011). Along with efforts to build up central authority, Afghanistan hosted important community-based initiatives for providing local infrastructure and services, as well as education specifically. In 2003, the central government established a community-driven development (CDD) initiative called the National Solidarity Program (NSP), with the support of the World Bank. The goal of NSP was to build basic infrastructure and rehabilitate villages throughout the country. Like other CDD initiatives, NSP used "radical decentralization" (Casey et al., 2018), giving communities direct control over decision-making and implementation processes in historically neglected parts of the country. In Afghanistan, NSP built off traditional shuras (village councils) and introduced new electoral processes to establish Community Development Councils in villages. The councils

were tasked with identifying local needs, submitting grant requests, and executing village-level infrastructure projects with the support of the Ministry of Rehabilitation and Rural Development. NSP, like other CDD programs, offered a successful institutional framework for addressing local infrastructural needs during the reconstruction process (Beath, Christia and Enikolopov, 2015).

Independent of NSP, community-based education (CBE) emerged as a dominant channel for education provision in remote areas, especially for girls who had been denied access to education under the Taliban (Burde and Linden, 2013; Burde, 2014). In contrast to NSP, CBE programs in Afghanistan were administered by external INGOs, who set up the classes or schools in existing village infrastructure like homes, mosques, or community buildings, using government textbooks, and recruited the teachers. Donors and organizations involved in CBE have conceptualized it as an "emergency" strategy for delivering education (Kelcey et al., 2019). The networks of CBE schools were managed centrally by the foreign-funded INGOs, with the ultimate intention being to transfer administration to the centralized government school network. CBE demonstrably improved childrens' access and learning in remote Afghan communities (Burde and Linden, 2013; Burde, Middleton and Samii, 2016). But the long-term goal of transitioning village classes into the centrally-administered government school network has been unsuccessful, as discussed in the introduction. Children lost access to education once the INGO programming ended because of failed handovers to the government (Burde, Middleton and Samii, 2016). Moreover, budgets for INGO-provided programs are often beyond what domestic governments claim that they can sustain; such was the view of the Afghanistan government with regard to CBE.

A question, then, is whether certain strengths from each of these community based strategies might be combined to create a sustainable service-delivery strategy. As it happens, the $3 billion Citizens' Charter National Priority Initiative (CC) launched by the office of the Afghan president in 2016 was established with precisely this idea in mind. The CC stated that all Afghans had a right to basic health and education. It proposed to make use of the

institutional network established by NSP and to connect it to high-quality INGO-run programs like CBE. Programs would have to operate at a budget level that could be sustained The "sustainability model" intervention studied in this paper was designed in partnership with the CC unit in the office of the presidency, the Afghan Ministries of Education, Finance and Reconstruction and Rural Development, and two INGOs. The goal was to operationalize and test the general logic of the CC initiative. The World Bank sponsored a parallel initiative during the same period to test the CC concept in health.

The study was completed prior to the Taliban seizing control of the government in summer 2021. This transition only increased both citizens' and onlookers' uncertainty about the stability and reach of any central government in Afghanistan. Arguably, this makes the analysis of decentralized approaches all the more salient.

4 Intervention

The Assessment of Learning Outcomes and Social Effects of Community-Based Education (ALSE) was a multi-year, mixed set of randomized controlled trials conducted in two phases to assess strategies for increasing and sustaining access to primary education in Afghanistan. The first phase of ALSE was carried out during an evaluation period from 2014 to 2016 and was designed to assess the effects of INGO-implemented CBE on boys' and girls' access to primary education and their level of learning (see Burde, Middleton and Samii, 2016). This first phase of ALSE provides an estimate of the impact of CBE under INGO administration relative to no CBE at all. These results are used below, but Phase Two is the focus in this paper. Phase Two tested the "sustainability model" based on the parameters of the CC policy. The goal was to increase the sustainability of INGO-initiated CBE classes after the withdrawal of INGOs. This second phase was carried out from 2016 to 2018 and put village-level Community Development Councils (CDCs) in charge of administering the CBE classes. Similar to community-based management interventions in other countries (Ganimian, 2016;

Patrinos, Barrera-Osorio and Fasih, 2009), the CDCs took on recruiting, paying, and monitoring teachers; facilitating the acquisition of space, school supplies, and other material needs; and attending to school-related problems as they arose.

The study proceeded cautiously, and prior to committing to transferring administrative responsibility to the CDCs, we conducted an assessment of the CDCs' institutional capacity. The capacity assessment used Bloom and Van Reenen (2007)'s methodology for measuring management practices, adapted for the context of education service administration (Bloom et al., 2015; Lemos and Scur, 2016). The assessment found that the CDC network offered a viable institutional infrastructure, but that some refresher training and training specific to school management would be required (Burde et al., 2017). We collaborated with the INGOs and Ministry of Education to design three- to five-day training workshops to address these needs and prepare the CDCs for taking over school management.

After the training was complete, administrative responsibilities were handed over from the INGOs to the CDCs, and CDCs administered the schools in their communities with the remote support of district and central Ministry of Education offices. The CDCs received funds on a monthly basis, via either electronic transfer or in-person visits either by CDC members to district capitals or through occasional visits by Ministry staff to the villages. These visits were far less frequent than the kinds of monitoring or assistance visits that INGOs conducted. The funds paid for teachers' salaries and school materials, and the CDCs could obtain guidance from Ministry officials as needed. The tasks that the CDCs had to execute were to ensure that the schools were well-provisioned and broadly accessible to students and that teachers were properly motivated to perform. As discussed above, the key risk was that these duties would not be implemented faithfully, potentially leaving students without access or materials or leaving teachers with no incentive to perform. The test of the sustainability model ran for a full academic year.

5 Randomized Equivalency Trial Design

Because the study was assessing a new governance strategy affecting a basic service in a context of high vulnerability, it was crucial to proceed in an exceptionally cautious manner.[2] Our formative research established the strong expectation that the status quo policy of handing INGO CBE classes off to centralized government administration would lead to their closure. The equipoise principle meant it would be ethically dubious to use that status quo policy as the control condition against which to compare the "sustainability model" (Wendler, 2021). Such concerns are especially relevant in a war-affected context with high vulnerability (Wolfe, 2020). Thus, we designed the study as a randomized "non-inferiority equivalency trial" (Lakens, Scheel and Isager, 2018) that tested the sustainability model against a gold standard benchmark of INGO administration. This section explains the research design.

The study took place in 132 villages situated in the predominantly rural provinces of Herat, Ghor, Bamiyan, Daykundi, Parwan, and Kapisa, which span the remote central highlands and western Afghanistan.[3] Candidate villages were randomly assigned to a treatment condition through a restricted randomization, whereby a treatment assignment profile was randomly selected from the set of treatment assignment profiles that satisfied a covariate balance criterion.[4] Specifically, 64 villages (which hosted 2,240 households and 2,688 children) were assigned to the treatment condition. The control condition was continued INGO administration in 68 villages (2,380 households and 2,856 children). To implement this, we secured grants from donors to extend the INGOs' CBE programming in the control villages. As discussed below, INGO-operated CBE classes ran at considerably higher cost than the sustainability model, and served as a high-performing benchmark against which to test the

[2] A pre-analysis plan was registered with the Evidence in Governance and Politics (EGAP) registry prior to receiving the outcome data.

[3] The broader study also included another 65 villages that were the control group for a prior phase of the research (discussed below). In the appendix, we describe the geographic context of the six provinces and discuss whether security conditions affected our ability to collect data (see Section A).

[4] Please refer to Table A6 in the Appendix for the covariate balance test results.

sustainability model, which was designed to operate under a budget constraint set by the Citizens' Charter. The objective was to assess whether the ALSE sustainability model is, at best, equivalent to INGO service provision and, at worst, only modestly inferior. In the spring of 2017, after the capacity assessments and just before the 1396 school year in the Afghan calendar, we worked with the Afghan MoE to facilitate the handover in the communities assigned to treatment.

This design avoided expected harms that would have been introduced by forcing control communities to operate under what we understood to be a dysfunctional status quo policy (Banerjee and Duflo, 2009; Humphreys and Weinstein, 2009; Teele et al., 2014). Of course, doing so sacrificed having a test of sustainability against the status quo handover policy. But this is also defensible in practical terms. With knowledge about the problems with the status quo policy, in actuality, the real choice at hand was whether we should resign to ongoing INGO administration or whether some strategy might exist to transfer administration to locally accountable authorities. The non-inferiority equivalence trial allows for comparing costs and benefits that inform this choice. Our approach is analogous to medical trials that examine whether a potentially less expensive experimental treatment (the sustainability model) may safely substitute for a currently-used expensive treatment (ongoing INGO administration).

Testing for equivalence is different than testing for a non-zero effect. A naive approach would be to compare to two treatment conditions, and then declare equivalence if we fail to reject the null hypothesis of zero effect. Such non-rejection of the null would perversely increase as statistical power decreased. We can address this problem by focusing attention on the lower bound of the confidence interval, which is the smallest effect that would be rejected at the given significant level. "Precisely-estimated zero" effects give stronger evidence of equivalence because the confidence intervals bounds are very close to zero. The question is how "precise" is precise enough? We can answer this question by comparing estimates to a threshold known as the "smallest effect size of interest" (SESOI) (Lakens, Scheel and

Isager, 2018; Rainey, 2014; Hartman and Hidalgo, 2018). This SESOI threshold defines a "worst-case" level of inferiority that would have us reject the experimental treatment in favor of the standard-of-care treatment. Essentially, one checks whether the lower bound of the confidence interval is above this worst-case threshold or not (see Figure 1).

We establish a SESOI threshold based on a worst-case trade-off between the sustainability model's lower cost and INGOs' potentially superior performance. It uses two inputs: first, an estimate of the gains from INGO administration relative to no CBE at all, which we label B, and second, a cost ratio that compares the cost of the experimental treatment (sustainability, labeled C_1) versus a standard-of-care control condition (INGO management, labeled C_0):

$$SESOI = -B \times (1 - C_1/C_0) \tag{1}$$

This SESOI expresses that we want to make sure that the effectiveness ratio is not worse than the cost ratio (see the Appendix for a derivation and an explanation for how this relates to Hendren and Sprung-Keyser (2022)'s "marginal value of public funds" criterion). This lower bound is justified by the imperative that the Afghan government faced under the Citizen's Charter initiative to reduce costs, but in a way that was not too detrimental to education.

Our judgment about the success of the sustainability model is based on an equivalency test for children's attendance and learning. We also use equivalency tests for other outcomes to assess potential mechanisms. To construct the SESOI for a given outcome, we use detailed cost accounting to determine C_1 and C_0. When possible, we use the average treatment effect (ATE) results from Phase One of the ALSE trial (described above) to establish a value for B. This applies to our main outcomes of attendance and standardized test scores. For outcomes that were not measured in Phase One but are included in Phase Two, we used a rule of thumb based on the observation that the Phase I ATEs for primary outcomes that were measured were equivalent to about 0.30 control group standard deviations, and so we used 0.30sd as the value for B in cases where we did not have ATE estimates from Phase I.

(SESOI calculations are discussed in greater detail in the Appendix.)

6 Data and Outcomes

Our data include two waves of post-treatment surveys with households, community leaders, and teachers, as well as children's learning assessments. The first wave was conducted between October and December 2017 at the conclusion of the academic year under study. All interviews were conducted face-to-face by trained enumerators from D3 Systems/ACSOR-Surveys, an independent survey firm. For the household surveys, the head of household or most senior person available was chosen for an interview and the team further collected roster data for other household members, including all children. Children present at the household and aged 6-13 were given a literacy and numeracy test based on Early Grade Reading Assessment and Math Assessment tools originally developed by RTI International and adapted by the research team to the Afghan context. Village heads of each study community and the teachers in each of the village schools were also interviewed. Using the American Association of Public Opinion Researchers' Response Rate 3 criterion, the final household response rate for the survey was 95.1%; for children on the learning assessments it was 78.4%; for community leaders it was 98.3%; and for teachers it was 98.6%. The second wave of data collection focuses on community leaders' longer-term sustainability perceptions. Our research team conducted a follow-up phone survey in summer 2018, a year after the conclusion of the intervention, with one community leader in each village in the ALSE sample. The realized distribution of clusters, villages, households and children in the treatment and control groups (after attrition) is detailed in the Appendix in Table A5. Survey questions and indicators for each of all outcomes are also outlined in the Appendix.

The primary *non-inferiority equivalency outcomes* are children's learning, children's attendance, and household satisfaction with education services. For each child in the household, the enumerators collected information on school access and attendance, as well as math and

literacy scores based on the learning assessment tests. The math and literacy outcome variable is created as a sum score, while the access and attendance outcome is constructed as a single binary variable. Household satisfaction is an index averaging a set of four indicators gauging household perception on access to quality education in the household survey.

A set of secondary outcomes captures potential agency problems. In particular, we collected information on the management and application of education funding (level of classroom supplies and regularity of teacher payments), teachers' intrinsic and extrinsic motivation and performance, and community leaders' own satisfaction with education in the community leader and teacher surveys. Each of these outcomes are similarly constructed as indices averaging across a set of indicators.

The third set of outcomes captures longer term sustainability perceptions. These variables are non-equivalency outcomes: we test the usual null hypothesis of whether perceptions are higher under the sustainability model versus under INGO administration. We collected information on community desire to continue CBE, the level of community engagement after the end of the ALSE intervention, and whether these CBE operations continued through spring 2018, a year after the ALSE intervention ended. The latter measure is simply a binary indicator of whether the CBE school was operational in spring 2018, while the former are indices constructed in a similar manner to the primary and secondary outcomes.

Summary statistics for outcomes are presented in Table 1. The sample size for the children (8-13) is smaller for the math and reading scores relative to the attendance outcome because the attendance sample is based on parents responses for the entire roster of children collected in the household surveys, while the math and reading scores are drawn from the learning assessment, which was administered to the children that were present. We tested for potential differential rates of missingness, and do not find meaningful differences between the treatment and control groups (Appendix Table A7). The sample size for the follow-up sustainability outcomes are considerably smaller than the first wave community leader survey sample size. The reason is due to the fact that one of the two INGOs in our study

continued their programming beyond the extra year that we funded, which then rendered our sustainability measurements irrelevant in these communities. As such, we only analyze outcomes in the communities serviced by either the sustainability model or the other INGO that concluded their program as originally expected. Because the design was stratified by INGO catchment area, the comparisons within the one INGO catchment areas is still internally valid, although we must be careful about generalizing to what might have obtained in the other INGO's catchment area.

The final column of Table 1 shows SESOI values for the equivalency outcomes, as calculated by equation (1). The cost ratio is based on village-level costs, since the biggest cost items are actually village-level fixed costs rather than student-level variable costs. We discuss the cost ratio calculation in the results section below. An "NA" in the last column means the outcome was used for a usual test of the null hypothesis of no difference, rather than an equivalency test.

Table 1: Summary Statistics for Equivalency and Non-Equivalency Outcomes

	N	Mean	St. Dev.	Min	Max	Outcome Type	SESOI
Children (Aged 8-13) Outcomes							
Attendance	4980	0.800	0.400	0.000	1.000	equivalency	-0.064
Math and reading score	3822	0.742	0.936	-1.602	1.827	equivalency	-0.131
Household Outcomes							
Household satisfaction	3754	-0.001	1.003	-3.548	1.012	equivalency	-0.139
Confidence in local institutions	3754	0.035	1.013	-2.893	1.185	non-equivalency	NA
Teacher attendance score	2545	-0.021	1.010	-2.827	0.463	equivalency	-0.139
Community perception of teacher quality	3754	-0.021	1.021	-3.560	1.732	equivalency	-0.139
Community Leader Outcomes							
Perceived access to quality education	129	-0.214	1.125	-3.322	0.800	equivalency	-0.139
Classroom supplies score	121	-0.058	1.216	-3.142	4.246	equivalency	-0.139
Combined sustainability index	121	0.023	0.903	-2.068	1.955	non-equivalency	NA
Community provision for future.classes	129	0.118	1.024	-0.978	3.089	non-equivalency	NA
Confidence in community institutions	121	-0.221	1.061	-3.308	2.114	non-equivalency	NA
Confidence that school will be sustained	121	0.011	1.003	-2.306	1.367	non-equivalency	NA
MOE provisions for funding teachers	129	0.001	0.913	-0.638	3.372	non-equivalency	NA
Shura provisions for future classes	129	0.143	1.019	-1.219	2.474	non-equivalency	NA
Teacher Outcomes							
Combined fund application index	154	0.253	0.958	-3.219	1.273	equivalency	-0.139
Classroom supplies	154	0.046	0.975	-4.095	0.761	equivalency	-0.139
Combined motivation index	154	0.066	0.945	-2.578	1.928	equivalency	-0.139
Intrinsic motivation	154	-0.024	0.955	-2.958	0.696	equivalency	-0.139
Material motivation and satisfaction	154	0.105	0.952	-3.233	2.235	equivalency	-0.139
Whether teacher is being paid regularly	154	0.308	0.937	-3.008	1.103	equivalency	-0.139
Combined sustainability index	154	-0.030	1.034	-2.030	3.078	non-equivalency	NA
Confidence in local community institutions	154	0.052	0.223	0.000	1.000	non-equivalency	NA
Confidence that school will be sustained	154	0.727	0.447	0.000	1.000	non-equivalency	NA
Proactive management from MOE	154	-0.115	0.955	-1.121	2.522	non-equivalency	NA
Proactive management from shuras	154	0.124	1.019	-1.576	3.600	non-equivalency	NA
Teacher plans to remain CBE teacher	154	-0.198	1.105	-2.733	0.618	non-equivalency	NA
Follow-up Sustainability Outcomes							
Community desire to continue CBE	61	0.984	0.128	0.000	1.000	equivalency	†
Combined sustainability index	60	0.011	1.270	-1.872	2.920	non-equivalency	NA
Initiative to have CBE sustained	61	0.361	1.277	-0.807	3.151	non-equivalency	NA
Provisions for sustainability without NGO	60	0.190	1.404	-1.368	4.568	non-equivalency	NA
Confidence that school will sustain	61	0.475	0.504	0.000	1.000	non-equivalency	NA
Continued CBE operations in spring.2018	61	0.311	0.467	0.000	1.000	non-equivalency	NA

Notes: †This variable has no variation in the control group, and so the SESOI cannot be defined.

7 Empirical Strategy

We implement the estimation strategy specified in our pre-analysis plan. We estimate average treatment effects and then test them against SESOIs for non-inferiority equivalency outcomes

and the usual null of zero for the non-equivalency outcomes. We use the following OLS regression specification:

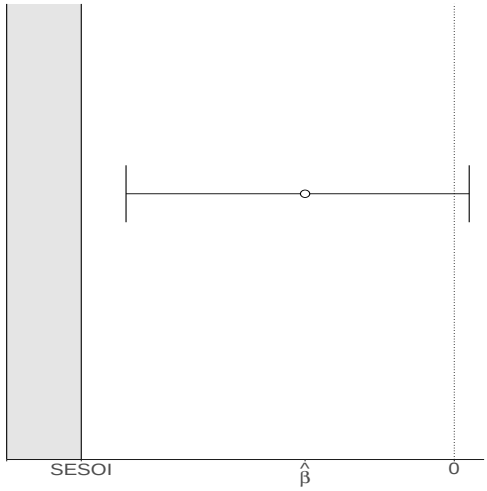
$$Y_{ics} = \alpha + \beta T_{cs} + X_{ics}\theta + \sigma_s + \epsilon_{ics} \tag{2}$$

for unit i (a household, child, community leader, or teacher) in village cluster c and randomization stratum s. The variable Y_{ics} is the outcome for unit i, T_{cs} is the treatment assigned to i's cluster within stratum s, X_{ics} is a vector of covariates, and σ_s characterizes the randomization-stratum heterogeneity. The treatment variable T_{cs} equals 1 for village clusters in the "sustainability model" treatment condition and 0 for those in the control condition of extended NGO administration. We use two covariate control vector X_{ics} specifications, as specified in our pre-analysis plan and as detailed in Table A4 in the Appendix. A longer list (specification cov2) contains a variety of household variables, and we use it on household and child assessment survey outcomes so as to maximize efficiency. A shorter list (specification cov1) has only basic community and respondent information, and we use it for the teacher and community leader surveys given that the sample sizes are one or two per community and therefore offer few degrees of freedom.

We used weighted least squares, where the weight accounts for the probability of being in the respective treatment/control condition and being sampled. We incorporate mean-centered randomization stratum indicators (corresponding to provinces) to partial out stratum-level heterogeneity, designated by σ_s in expression (2). To account for the fact that treatment assignment was at the cluster level, we construct finite-sample adjusted cluster-robust confidence intervals based on Pustejovsky and Tipton (2018). We target an error rate of $\alpha = 0.05$.

Figure 1 illustrates the logic of our non-inferiority equivalency tests. For equivalency tests, we are interested in one-sided inference to determine the levels of inferiority that we can reject with 95% confidence. For a one-sided test at 95% confidence, the relevant quantity is the lower bound on a 90% confidence interval (Lakens, Scheel and Isager, 2018; Rainey, 2014). The SESOI provides us with the benchmark for the test. If the 90% confidence interval of our estimated ATEs ($\hat{\beta}$) is not fully above the SESOI, then we fail to reject the hypothesis

Figure 1: The Logic of a Test for Non-Inferiority



Notes: Even though the point estimate is negative and the confidence interval overlaps with zero, the interval is fully above the SESOI. Thus, in this example, we would conclude that the experimental treatment exceeds the threshold needed to recommend it over the more costly control condition.

that the sustainability model is "unacceptably inferior" to INGO administration. If the 90% interval is fully above the SESOI, we reject the hypothesis of unacceptable inferiority and conclude that the experimental treatment exceeds the minimum performance benchmark. In Figure 1, $\hat{\beta}$ represents an estimate of the effect of the sustainability model relative to extended INGO administration. The upper and lower bounds of the confidence interval for this estimate are marked on the plot. The interval is well above the SESOI. We would conclude that the sustainability model exceeds the minimal performance benchmark and can therefore be recommended over the control condition. For the non-equivalency hypotheses, we report the usual two-sided 95% confidence intervals or significance tests.

8 Main Results

8.1 Cost Comparison

We first report on the cost difference, which we calculated using a detailed ex post audit. (See Table A2 in the Appendix.) The cost comparison is with respect to *ongoing* yearly costs, and therefore omits fixed startup costs that the INGOs took on to launch CBE in a village. We calculated ongoing costs for the INGOs using multiple years of data, and then calculated a comparable amount for the sustainability model through a careful accounting of ongoing costs. We calculated an average per-village yearly cost of USD 6,439 under INGO management (C_0) versus USD 3,457 under the sustainability model (C_1). This amounts to a cost ratio of 0.537, which is what we use in establishing the SESOI for our non-inferiority equivalency tests. Such a cost savings meets the demands of the Citizen's Charter initiative.

Looking at the details, the keys to the cost difference comes from INGOs' very high relative expenditures for monitoring the schools and administrative staff costs. Teacher salaries paid out were slightly higher under the sustainability model (USD 1202.85 versus USD 932.66). But the INGOs spent USD 1143 per village on monitoring versus USD 236 under the sustainability model, nearly a five-fold difference. This aligns perfectly with our theoretical framework: under community administration, schools are monitored by leaders within the community; so long as local leaders are kept honest, their local monitoring effectively replaces monitoring by bureaucrats who have to regularly travel out from the center. Administrative staff costs (covering the project coordinators based at the MoE for the sustainabilty model and INGO staff under INGO administration) were substantially lower under the sustainabilituy model as well (USD 1,078.78 versus USD 2,641.67 on a per village basis).

The difference in staff costs raises an important issue of uncompensated labor by community leaders and councils under the sustainability model. When engaging community leaders and councils to participate in the test of the sustainability model, it was clarified that the work would be non-compensated for the period of the study, and communities' consent was

sought on that basis. That consent was universal speaks to leaders' and council members' willingness to contribute on an in-kind basis. Nonetheless, effort put into managing school takes away from other things, and having community members do this unpaid may not be sustainable in the long run (Alatas et al., 2012; Casey et al., 2018). Without time-use data, we cannot calculate these hidden costs directly. What we do below is to derive the implications for how much community councils could be paid without sacrificing the attractiveness of the sustainability model relative to INGO administration.

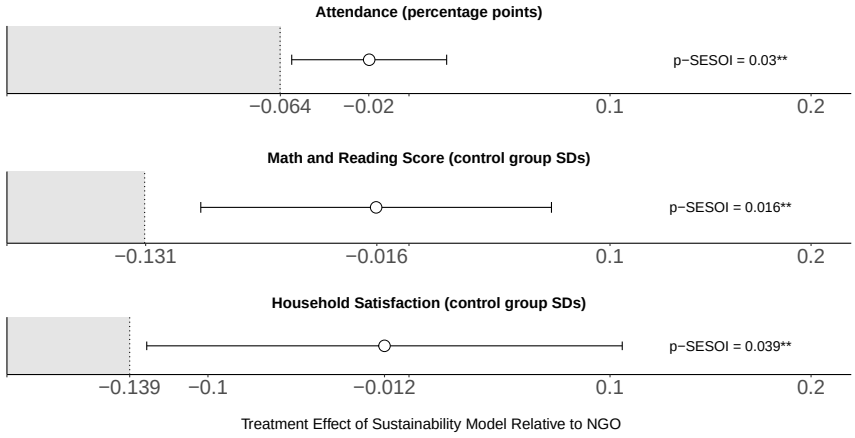
8.2 Primary Equivalence Results

Figure 2 presents the results of the primary non-inferiority equivalency tests for attendance, learning, and household satisfaction. We do not expect the sustainability model to outperform INGO administration, but we want to see whether it clears a benchmark that distinguishes acceptable from absolutely unacceptable performance. Overall, the results indicate that the sustainability model performs well. It exceeds the cost-adjusted benchmark. While school attendance and access under the sustainability model is 2 percentage points lower than under INGO administration, what we want to know is whether it would be less than the SESOI of -6.4 percentage points. We reject this, with a p-value of 0.03.

The middle panel shows effects for math and reading scores. The estimated difference is small, with the sustainability model performing 0.016 control group standard deviation units lower than INGO administration. This is substantially above the SESOI of -0.131, implying that the sustainability model well exceeds the benchmark.

In the bottom panel, we show results for household satisfaction with education services. The outcome is an index composed of indicators of household perceptions of access to quality education. The index is scaled in terms of control group standard deviations. Household satisfaction is only 0.013 control group standard deviations lower than under NGO administration. This is significantly above the SESOI of -0.139, meaning that satisfaction levels exceed the benchmark.

Figure 2: Primary Equivalence Results



Notes: Treatment effects with interval to test against SESOI with $\alpha = 0.05$. The p values are for tests against the SESOI, with ** significant at $\alpha = 0.05$. The gray area distinguishes between areas above (white) and below (gray) the SESOI, which is the benchmark that we use to determine whether the sustainability model surpasses minimum performance standards given the cost ratio. For all three of these primary outcomes, the sustainability model surpasses this performance standard with p-values less than 0.05.

Turning to the question of compensating community leaders and councils, the lower bound on the 95% confidence interval for the difference in attendance rates is -0.058. Using the SESOI expression, this would imply that at least $250 could be added to the per-village cost of the sustainability model and still have it be preferable to INGO implementation with 95% confidence. Per current International Labor Organization figures, this is double the average full-time monthly wage in Afghanistan in 2020.[5] For math and reading scores, the lower bound on the 95% confidence interval is -0.104, which still leaves the possibility of adding almost $700 per year to the cost. Thus, our conclusions do not depend on ignoring the labor of the community leaders and councils.

[5]As per data posted to https://ilostat.ilo.org/data/country-profiles/ (accessed January 2022). Inflation figures from World Bank data posted to https://data.worldbank.org/ (accessed January 2022) suggest current 2020 wages Afghanistan were about 10% higher than during the study period.

8.3 Effect Heterogeneity by Gender and Ethnicity

We now check whether community-managed education leads to inequities due to discrimination against girls or ethnic minorities. To examine this, we estimate effects conditional on gender and ethnic status. In particular, we defined ethnic status in two ways. First, we created a binary indicator for whether a child is part of an ethnic majority or minority group within their village and, secondly, we coded a binary measure for whether a child shared the same ethnic group as the community leader in his or her village. We conduct Wald tests for effect heterogeneity between genders and ethnic status categories in the effects of the sustainability model. For one-sided tests of whether girls fair worse, with $\alpha = 0.05$, our sample sizes provide power to detect effect differences of 8 percentage points for attendance and around 0.19 standard deviations for test scores. For tests of whether ethnic minorities fair worse, we have power to detect differences of about 7-8 percentage points for attendance and 0.20-0.25 standard deviations for test scores. At this level, we find no significant differences in effects by gender or ethnic status (Table A8 in the Appendix), alleviating concerns about substantial gender or ethnic biases in implementation.

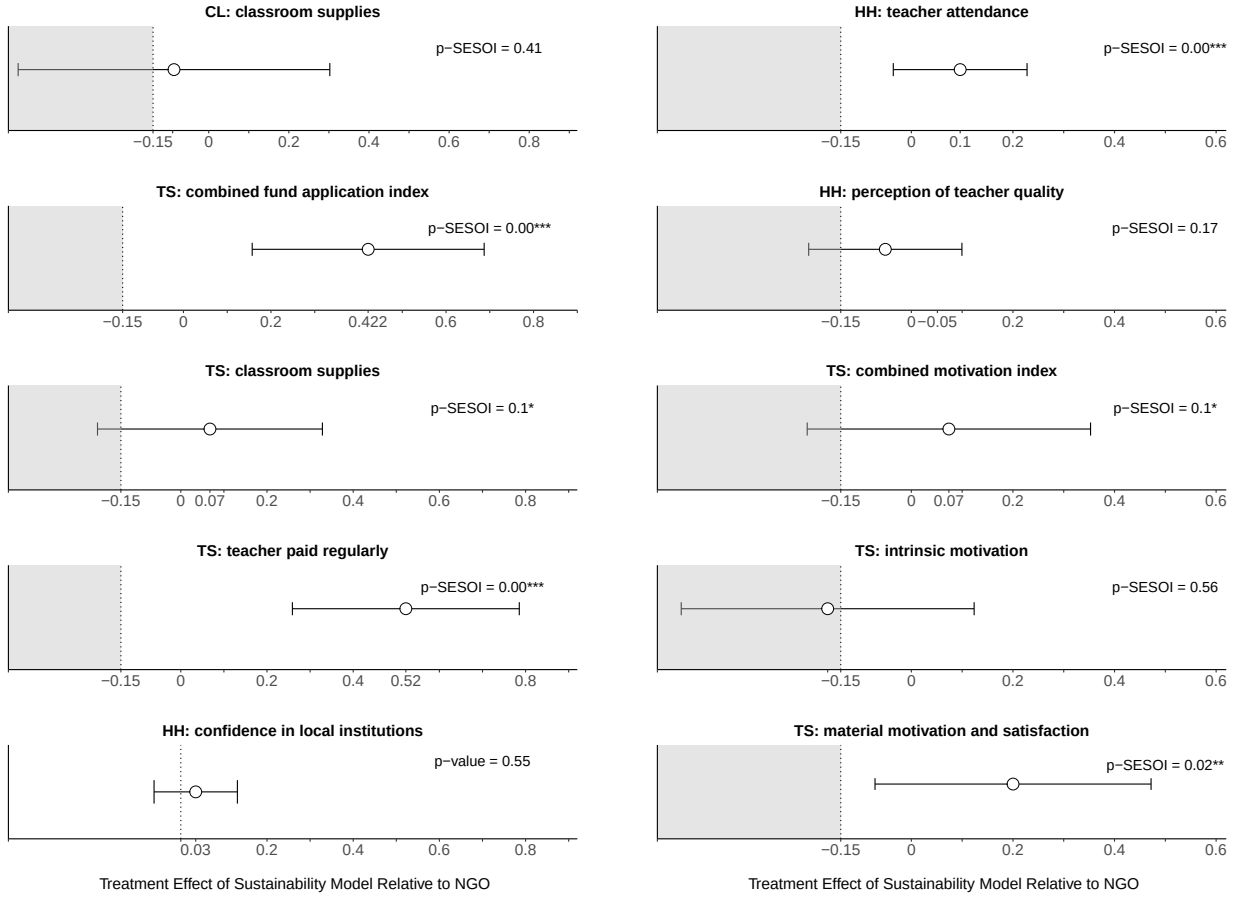
8.4 Checking for Agency Problems

As we discussed in Section 2, delegation to community institutions could increase resource misappropriation or other agency problems. Such risks may be mitigated by local accountability or the intrinsic interest of community leaders in having their community's (and their own) children educated. In this section, we show that agency problems did not undermine the sustainability model and that, instead, it performs better than INGO administration by some measures. We collected information in the community leader, teacher, and household surveys on use of funds. We created indices measuring the level of classroom supplies and whether teachers were paid regularly, and we combined these into a fund application index. Each of these outcomes was tested under non-inferiority equivalency hypotheses. To examine potential concerns over misuse of resources, we collected information on households'

confidence in local institutions, measured under the non-equivalency hypothesis.

Panel A of Figure 3 presents these results. For the classroom supplies measure from the community leader survey, we cannot rule out the possibility that the sustainability model had an unacceptably large negative effect (estimates equals -0.09 sd units) on classroom supplies, although statistical power is very low for this test. However, from the teacher's survey, it appears that the sustainability model met the performance benchmark in terms of availability of supplies and payment of salaries. Indeed, it seems that teachers had better access to funds and were paid *more* regularly under the community-managed sustainability model than under INGO administration, in line with findings in Bold et al. (2018). Further, household confidence in local institutions increases by 0.03 sd units relative to INGOs, but this effect is not statistically significant at conventional levels.

Figure 3: Test of Potential Agency Issues



(a) Education Funding and Confidence (b) Teacher Motivation and Performance

Note: Treatment effects with interval to test against SESOI for equivalency outcomes and against 0 for non-equivalency outcomes, with $\alpha = 0.05$. * significant at $\alpha = 0.10$, ** significant at $\alpha = 0.05$, *** significant at $\alpha = 0.01$.

We also examine if there may be intra-community agency problems in being able to keep teachers motivated under the sustainability model relative to INGO teachers. We measure this using household reports on teacher attendance and perceptions of teacher quality, as well as surveying the teachers themselves to evaluate their own intrinsic and extrinsic motivation for teaching the CBE classes. All of these outcomes are tested under equivalency hypotheses and are presented in Panel B of Figure 3. While household results suggest that teacher attendance met the cost-adjusted performance benchmark, this was not quite the case for perceptions of teacher quality. Teachers report higher overall motivation, an effect driven

by a 0.20 sd unit increase in their material motivation and satisfaction under the sustainability model, although this is contrasted with a considerable decline, though statistically insignificant, in reported intrinsic motivation. These findings are a mix of strongly positive and inconclusive results when it comes to intra-community agency issues.

In our theoretical discussion we proposed that leaders could have either extrinsic motivation, induced by their constituents' demand for education, or intrinsic motivation to carry out their administrative responsibilities faithfully. We explore this question by examining heterogeneous treatment effects based on whether the community leader in a village has school-aged (8-13) children or not. We look at children's attendance and learning outcomes, household satisfaction and confidence in local institutions outcomes, and household evaluations of teacher attendance and quality (in Table A9 in the appendix). We find no statistically significant differences. The sustainability model does not depend crucially on community leaders' *personal* interest in the schools. We recall that households across the board put priority on education, and so this variation in leaders' personal circumstances may be irrelevant given constituents' interests for leaders to deliver on education.

8.5 Checking for Lingering Capacity Deficits

The sustainability model was deployed following a systematic assessment of the capacity of the community councils and their leaders. We used this assessment to design a training program to prepare the councils for CBE administration. We evaluate whether the training succeeded in leveling out capacities. We test for whether pre-training capacity levels moderate the intervention's effects. The results are show in the appendix in Tables A10-A12, which, respectively, show interaction (i.e. moderator) effects based on a sum score for measures of general management capacity, a principal components score for indicators of previous training, and a principal components score for indicators of whether the council was active at the time of the assessment.[6] Looking at general management capacities in Table A10,

[6]For additional details on the capacity assessment, see Burde et al. (2017).

the results show that for many of the outcomes, localities with higher capacity councils performed better under *INGO adminstration*. For example, this holds rather strongly for household perceptions of teacher attendance and community leader (CL) assessments of how well classes were funded using the first two capacity measures. In these places, the councils do not have a formal role, but these patterns may indicate the councils playing a role in the INGO's work. The positive correlation disappears under the sustainability model (the interaction term coefficients are of the opposite sign and similar magnitude of the coefficient on the capacity scores). Such findings are consistent with the goal of the capacity training, which was to ensure that all communities selected to implement the sustainability model were prepared to do so.

9 Leaders' and Teachers' Perceptions of Long Term Sustainability

The results above show that the sustainability model met our primary performance benchmarks relative to INGO administration. The question remains about how the sustainability model might help to secure access to primary education in the long run. Our study funded the sustainability model and collaborated with government ministries to ensure higher level institutional supports for a limited duration as a trial. The results of the trial (as presented here) would inform higher-level policy decisions about institutionalizing it (or not). In the meantime, we collected data both within the original survey and in a one-year follow-up to help us assess what might be needed to institutionalize the sustainability model.

The first set of results show that the sustainability model did not make community leaders and teachers more confident about long-term access to schooling. In the endline surveys run at the end of the trial period, we asked community leaders and teachers about whether they thought schooling arrangements in their communities were sustainable. We also asked community leaders about Ministry of Education and community council provisions for

future classes and funding. We asked teachers about proactive management by community councils and government institutions, as well as the teachers' plans to remain as teachers in their villages. We constructed combined sustainability indices for teachers and community leaders, averaging these component indicators. Effects on these outcomes were tested under non-equivalency hypotheses to evaluate whether the sustainability model improved these outcomes relative to INGO provision or not (refer to Appendix Table A13). We mostly find insignificant effects and sometimes negative effects. The aggregate (index) effect for community leaders was statistically insignificant, reflecting a mix of positive and negative effects on the component indicators (Panel A). Teachers were more pessimistic, and the effect on the index is substantial (-0.45sd), reflecting lower confidence that schools will be sustained and, perhaps consequently, that they will remain as a CBE teacher. These results show that experience of community administration of CBE did not, in itself, inspire confidence that schools could be sustained beyond the trial period.

We explored what might drive such low confidence. Community leaders and teachers presumably understood that the government had not yet established a way to keep the sustainability model going beyond the trial, leading to uncertainty. Community leaders may have also doubted their own capacity to administer schools without INGO help or more government support. Teachers may have doubted community leaders' capacities as well. To evaluate this possibility, we tested for how leaders' age (above or below the median age of 52) and level of education (any formal education versus no education) affected their perceptions. The results are shown in Appendix section K. We estimate some substantial moderator effects with respect to age: pessimism with regard to community provisions for classes and confidence in community institutions is concentrated among community leaders above the median age of 52. We also explored whether there was any association between how well children scored in the math and reading tests and leaders' perceptions. This was to check whether leaders' perceptions were driven by what they may have witnessed in terms of school performance. These results are shown in Appendix Table A17. We

find little indication of such possibilities. For many outcomes, the control-group (INGO community) correlation between test scores and sustainability outcomes appears as negative, although the estimates are too noisy in most cases to reject the null (the exception is for confidence in community institutions). To the extent that such correlations exist in the control condition, they are canceled out by the sustainability model treatment. Thus, in the sustainability model communities there is essentially zero correlation between communities' mean test scores and community leaders' sustainability perceptions. The findings suggest that, if anything, leaders ought to be reassured that students did well in the schools that they helped to administer, and that they should feel confident in their capacity to administer. Older leaders may especially need encouragement or additional supports to feel confident in leading the sustainability model.

For the longer term follow-up, statistical power limits our ability to draw strong conclusions. The longer term follow-up was done with a phone survey with community leaders in spring 2018, about a year after the endline survey. Our analysis is confined to communities that fall within the catchment area of only one of the two INGOs with whom we collaborated. Our initial design was based on the understanding that, after 2018, all classes in INGO-administered communities would be handed over to the MoE for government administration as per the status quo handover policy. However, one of the INGOs unexpectedly decided not to hand over their classes, but rather raised new funds to continue running their schools. The villages covered by this INGO did not offer a sensible comparison, since no handover occurred. Thus, we limit our comparison to the communities where status quo handover occurred as expected. This offers a valid comparison but low statistical power. (We present the results for the entire sample and address the generalizability questions in Appendix sections I and J.) We asked leaders to rate their communities' desire to continue with the village schools, initiative undertaken and provisions in place to sustain CBE, confidence that schools will sustain in the long-run, and then whether the school was operational at the time (tested under non-equivalency hypotheses). Table 2 shows that leaders rated

their community's desire to sustain schools about the same in the sustainability model and the INGO communities.[7] Community initiative to sustain CBE is higher after the sustainability model (0.48sd higher), although this is not statistically significant. Schools continued to operate in 31% of villages that had been serviced by INGOs and a nearly identical 28% of sustainability model villages. This would have occurred through the government taking on administrative responsibility or communities keeping the schools going on their own. Putting all of the outcomes together in the sustainability index, we do not obtain a clear finding of either harm or benefit (we estimate a 0.44sd increase, but this is not statistically significant). We reiterate though that these results were obtained prior to the government having made any policy changes to adopt the sustainability model.

Table 2: Community Leaders' Longer-term Perceptions of CBE Sustainability (1 Yr. Follow-up)

	Treatment Effect	Std. Err.	N	Control Mean	Control Std. Dev.
Equivalency Outcome					
Community desire to continue CBE†	-0.04	(0.04)	61	0.98	0.13
Non-Equivalency Outcomes					
Combined sustainability index	0.44	(0.31)	60	0.01	1.27
Initiative to have CBE sustained	0.48	(0.31)	61	0.36	1.28
Provisions for sustainability without NGO	0.30	(0.34)	60	0.19	1.40
Confidence that school will sustain	0.05	(0.13)	61	0.48	0.50
Continued CBE operations in spring 2018	-0.03	(0.12)	61	0.31	0.47

Note: CR2 cluster robust standard errors shown in parentheses. All regression use the "cov2" specification.
† This outcome had no variation in the control group, and so a standard-deviation based SESOI could not be defined. In treatment there was one observation with a 0, and in control all observations had a value of 1.

[7] In our pre-analysis plan, we specified that we would formally test this as an equivalency hypothesis. However, because there is zero variation in the control group, a SESOI measured in terms of control group standard deviations is undefined. We thus evaluate the treatment effect point estimate in terms of the substantive significance.

10 Conclusion

We find that when higher level institutional supports are in place, decentralized community-based education administration in Afghanistan can sustain what INGO schools achieve in promoting childrens' access to education, their learning, and households' satisfaction with schooling. The community-based sustainability model significantly surpassed expectations given considerable cost differences relative to INGO administration. This shows that a model of service administration that is rooted in mechanisms of bottom-up accountability can perform in a setting in which top-down, centralized approaches have previously failed.

We find no evidence of significant discrimination, misdirection of resources, or negative impacts on teacher performance under the sustainability model relative to INGO management. The sustainability model performed well enough so as to allow for community leaders to be compensated at a reasonable level, while still maintaining a cost-effectiveness advantage over ongoing centralized INGO management. Given how unfavorable the broader institutional conditions were in Afghanistan, we have good reason to believe that these findings would be robust to a wide variety of contexts.

These primary findings point to the promise of a community-based sustainability model as a viable alternative to the status quo of handover to failure-prone top-down centralized government administration in places like Afghanistan. Such positive examples are important for "scale-up" initiatives, such as USAID's Development Innovation Ventures and the multilateral Global Innovation Fund. It is also worth reiterating that the context was about as unfavorable as one could imagine in terms of complementary institutions. While the sustainability model should not be viewed as a substitute for INGOs in *initiating* programs, it does help to ensure that services do not disappear after INGOs depart. INGOs can contribute to local community capacities to work with central government as part of the statebuilding process, complementing other strategies for enhancing governance in fragile states (Child, Wright and Xiao, 2021; Lyall, Zhou and Imai, 2020; Berman et al., 2019).

Nonetheless, our analysis of community leaders' and teachers' confidence in whether

CBE classes would continue under the sustainability model, as well as follow up surveys with leaders one year after the conclusion of the pilot, reveal mixed results. Without clear commitments of ongoing support for the model, these communities faced uncertainty after the trial period. Thus the positive findings here must be understood as coming in the context of exogenous intervention to ensure higher level institutional supports. These results are in line with what Bold et al. (2018) and Banerjee et al. (2017) discuss as the importance of complementary policies that deal with implementation constraints associated with sustaining reforms.

This paper provides a new take on "community directed development" (CDD). We emphasize CDD bottom-up accountability mechanisms and the combination of induced and intrinsic interests among local elites as ways to overcome agency problems in decentralizing service administration. The findings are not quite vindication of CDD, because we do not have a non-CDD comparison in our study. Nonetheless, it does show how CDD investments can produce a durable institutional network that can serve basic needs, so long as higher level instititional supports are in place. The findings here support Mansuri and Rao (2012)'s "sandwich model."

While the sustainability model studied here allowed for low cost education provision without glaring implementation issues, there are a number of remaining questions and concerns that this paper opens up. In particular, in spite of the success of this decentralized approach, challenges remain with regards to the adaptability and willingness of centralized bureaucracies to take up this new model for education provision. Furthermore, while the sustainability model examines the feasibility for communities to sustain what an INGO has *already* set up, our research design does not allow for assessing the sustainability implications or possible agency problems that may arise when community institutions manage education, or another public service, without any existing infrastructure set up by an INGO. We leave the exploration of these possible political consequences to future research.

References

Alatas, Vivi, Abhijit Banerjee, Rema Hanna, Benjamin A Olken and Julia Tobias. 2012. "Targeting the poor: evidence from a field experiment in Indonesia." *American Economic Review* 102(4):1206–40.

Alatas, Vivi, Abhijit Banerjee, Rema Hanna, Benjamin A Olken, Ririn Purnamasari and Matthew Wai-Poi. 2019. Does elite capture matter? Local elites and targeted welfare programs in Indonesia. In *AEA Papers and Proceedings*. Vol. 109 pp. 334–39.

Alesina, Alberto and Guido Tabellini. 2007. "Bureaucrats or politicians? Part I: a single policy task." *American Economic Review* 97(1):169–179.

Altschuler, Daniel. 2013. "How patronage politics undermines parental participation and accountability: Community-managed schools in Honduras and Guatemala." *Comparative Education Review* 57(1):117–144.

Andrabi, Tahir, Jishnu Das, Asim Ijaz Khwaja, Tara Vishwanath and Tristan Zajonc. 2008. *Pakistan Learning and Educational Achievement in Punjab Schools (LEAPS) - Insights to inform the education policy debate*. Washington, DC: World Bank.

Bandiera, Oriana, Michael Callen, Katherine Casey, Eliana La, Camille Landais Ferrara and Matthieu Teachout. 2019. "State effectiveness." *IGC Evidence Paper* .

Banerjee, Abhijit, Rukmini Banerji, James Berry, Esther Duflo, Harini Kannan, Shobhini Mukerji, Marc Shotland and Michael Walton. 2017. "From proof of concept to scalable policies: Challenges and solutions, with an application." *Journal of Economic Perspectives* 31(4):73–102.

Banerjee, Abhijit V and Esther Duflo. 2009. "The experimental approach to development economics." *Annual Review of Economics* 1(1):151–178.

Banks, Nicola, David Hulme and Michael Edwards. 2015. "NGOs, states, and donors revisited: Still too close for comfort?" *World development* 66:707–718.

Bardhan, Pranab. 2002. "Decentralization of governance and development." *Journal of Economic perspectives* 16(4):185–205.

Bardhan, Pranab and Dilip Mookherjee. 2005. "Decentralizing antipoverty program delivery in developing countries." *Journal of public economics* 89(4):675–704.

Bardhan, Pranab and Dilip Mookherjee. 2006. "Decentralisation and accountability in infrastructure delivery in developing countries." *The Economic Journal* 116(508):101–127.

Beath, Andrew, Fotini Christia and Ruben Enikolopov. 2015. *The national solidarity program: Assessing the effects of community-driven development in Afghanistan*. The World Bank.

Beath, Andrew, Fotini Christia and Ruben Enikolopov. 2017. "Direct democracy and resource allocation: Experimental evidence from Afghanistan." *Journal of Development Economics* 124:199–213.

Beath, Andrew, Fotini Christia and Ruben Enikolopov. 2018. "Do elected councils improve governance? Experimental evidence on local institutions in Afghanistan." *Technical Report, MIT Political Science Department* .

Berman, Eli, Michael Callen, Clark C Gibson, James D Long and Arman Rezaee. 2019. "Election fairness and government legitimacy in Afghanistan." *Journal of Economic Behavior & Organization* 168:292–317.

Berman, Eli, Michael Callen, Luke N Condra, Mitch Downey, Tarek Ghani and Mohammad Isaqzadeh. 2017. "Community Monitors vs. Leakage: Experimental Evidence from Afghanistan." *Technical Report, University of California-San Diego* .

Björkman, Martina and Jakob Svensson. 2009. "Power to the people: evidence from a randomized field experiment on community-based monitoring in Uganda." *The Quarterly Journal of Economics* 124(2):735–769.

Blair, Robert A and Matthew S Winters. 2020. "Foreign aid and state-society relations: Theory, evidence, and new directions for research." *Studies in Comparative International Development* 55(2):123–142.

Bloom, Nicholas and John Van Reenen. 2007. "Measuring and explaining management practices across firms and countries." *The quarterly journal of Economics* 122(4):1351–1408.

Bloom, Nicholas, Renata Lemos, Raffaella Sadun and John Van Reenen. 2015. "Does management matter in schools?" *The Economic Journal* 125(584):647–674.

Bold, Tessa, Mwangi Kimenyi, Germano Mwabu, Justin Sandefur et al. 2018. "Experimental evidence on scaling up education reforms in Kenya." *Journal of Public Economics* 168:1–20.

Burde, Dana. 2014. *Schools for Conflict or for Peace in Afghanistan*. Columbia University Press.

Burde, Dana, Joel A Middleton and Cyrus Samii. 2016. *The Assessment of Learning Outcomes and Social Effects of Community-Based Education: A Randomized Field Experiment in Afghanistan, Phase One Outcomes Report*. Steinhardt School of Culture, Education, and Human Development, New York University.

Burde, Dana and Leigh L Linden. 2013. "Bringing education to Afghan girls: A randomized controlled trial of village-based schools." *American Economic Journal: Applied Economics* 5(3):27–40.

Burde, Dana, Matthew Lisiecki, Joel Middleton, Otgonjargal Okhidoi, Mohammad Amin Sadiqi, Cornelia Sage and Cyrus Samii. 2017. *The Assessment of Learning Outcomes and Social Effects of Community-Based Education: A Randomized Field Experiment in Afghanistan, Phase Two Baseline Report*. Steinhardt School of Culture, Education, and Human Development, New York University.

Burde, Dana, NYU Joel Middleton, Matthew Lisiecki, Joel Middleton, Otgonjargal Okhidoi and Cyrus Samii. 2019. *The Assessment of Learning Outcomes and Social Effects of Community-Based Education: A Randomized Field Experiment in Afghanistan, Phase Two Endline Report*. Steinhardt School of Culture, Education, and Human Development, New York University.

Campbell, Susanna, Matthew DiGiuseppe and Amanda Murdie. 2019. "International development NGOs and bureaucratic capacity: Facilitator or destroyer?" *Political Research Quarterly* 72(1):3–18.

Casey, Katherine, Rachel Glennerster, Edward Miguel and Maarten Voors. 2018. "Skill versus voice in local development." *The Review of Economics and Statistics* pp. 1–46.

Chattopadhyay, Raghabendra and Esther Duflo. 2004. "Women as policy makers: Evidence from a randomized policy experiment in India." *Econometrica* 72(5):1409–1443.

Child, Travers, Austin L Wright and Yun Xiao. 2021. "Aid Fragmentation and Corruption." *University of Chicago, Becker Friedman Institute for Economics Working Paper* (2021-69).

Christia, Fotini. 2019. Decentralization in post-conflict settings: Assessing community-driven development in the wake of violence. In *Decentralized governance and accountability*. Cambridge University Press pp. 205–228.

Congressional Research Service. 2020. *The Washington Post's Afghanistan Papers and U.S. Policy: Main Points and Possible Questions for Congress*. Congressional Research Service.

Dasgupta, Aniruddha and Victoria A Beard. 2007. "Community driven development, collective action and elite capture in Indonesia." *Development and change* 38(2):229–249.

De Regt, Jacomina, Shruti Majumdar and Janmejay Singh. 2013. *Designing community-driven development operations in fragile and conflict-affected situations: Lessons from a stocktaking*. The World Bank.

Deserranno, Erika, Aisha Nansamba and Nancy Qian. 2020. "The unintended consequences of NGO-provided aid on government services in Uganda." *NBER Working Paper* (w26928).

Duflo, Esther, Pascaline Dupas and Michael Kremer. 2015. "School governance, teacher incentives, and pupil–teacher ratios: Experimental evidence from Kenyan primary schools." *Journal of public Economics* 123(2015):92–110.

Edwards, Michael and David Hulme. 1996. "Too close for comfort? The impact of official aid on nongovernmental organizations." *World development* 24(6):961–973.

Ensminger, Jean. 2017. *Corruption in community-driven development. A Kenyan case study with insights from Indonesia*. Chr. Michelsen Institute.

Faguet, Jean-Paul. 2004. "Does decentralization increase government responsiveness to local needs?: Evidence from Bolivia." *Journal of public economics* 88(3-4):867–893.

Fox, Jonathan A. 2015. "Social accountability: what does the evidence really say?" *World Development* 72:346–361.

Fritzen, Scott A. 2007. "Can the design of community-driven development reduce the risk of elite capture? Evidence from Indonesia." *World Development* 35(8):1359–1375.

Gadenne, Lucie and Monica Singhal. 2014. "Decentralization in developing economies." *Annu. Rev. Econ.* 6(1):581–604.

Galasso, Emanuela and Martin Ravallion. 2005. "Decentralized targeting of an antipoverty program." *Journal of Public economics* 89(4):705–727.

Galiani, Sebastian, Paul Gertler and Ernesto Schargrodsky. 2008. "School decentralization: Helping the good get better, but leaving the poor behind." *Journal of public economics* 92(10-11):2106–2120.

Ganimian, Alejandro J. 2016. "Why do some school-based management reforms survive while others are reversed? The cases of Honduras and Guatemala." *International Journal of Educational Development* 47:33–46.

Hartman, Erin and F Daniel Hidalgo. 2018. "An equivalence approach to balance and placebo tests." *American Journal of Political Science* 62(4):1000–1013.

Hendren, Nathaniel and Ben Sprung-Keyser. 2022. "The Case for Using the MVPF in Empirical Welfare Analysis." *National Bureau of Economic Research* .

Humphreys, Macartan and Jeremy M Weinstein. 2009. "Field experiments and the political economy of development." *Annual Review of Political Science* 12:367–378.

Jimenez, Emmanuel and Yasuyuki Sawada. 1999. "Do community-managed schools work? An evaluation of El Salvador's EDUCO program." *The world bank economic review* 13(3):415–441.

Jochem, Torsten, Ilia Murtazashvili and Jennifer Murtazashvili. 2016. "Establishing local government in fragile states: Experimental evidence from Afghanistan." *World Development* 77:293–310.

Kelcey, Jo, Amy Kapit, Dana Burde and Cyrus Samii. 2019. "Community Based Education and the Struggle to Support Post-Conflict Statebuilding in Afghanistan: Lessons for Non-State Service Providers." *Technical Report, Steinhardt School, New York University* .

King, Elisabeth and Cyrus Samii. 2014. "Fast-track institution building in conflict-affected countries? Insights from recent field experiments." *World Development* 64:740–754.

Lakens, Daniël, Anne M Scheel and Peder M Isager. 2018. "Equivalence testing for psychological research: A tutorial." *Advances in Methods and Practices in Psychological Science* 1(2):259–269.

Lemos, Renata and Daniela Scur. 2016. "Developing management: An expanded evaluation tool for developing countries." *Technical Report, London School of Economics, Centre for Economic Performance, London* .

Lund, Jens Friis and Moeko Saito-Jensen. 2013. "Revisiting the issue of elite capture of participatory initiatives." *World development* 46:104–112.

Lyall, Jason, Yang-Yang Zhou and Kosuke Imai. 2020. "Can economic assistance shape combatant support in wartime? Experimental evidence from Afghanistan." *American Political Science Review* 114(1):126–143.

Mansuri, Ghazala and Vijayendra Rao. 2004. "Community-based and-driven development: A critical review." *The World Bank Research Observer* 19(1):1–39.

Mansuri, Ghazala and Vijayendra Rao. 2012. *Localizing development: Does participation work?* World Bank Publications.

Murtazashvili, Jennifer Brick. 2016. *Informal order and the state in Afghanistan*. Cambridge University Press.

Murtazashvili, Jennifer and Ilia Murtazashvili. 2020. Bottom-Up State-Building. In *Bottom-up Responses to Crisis*. Springer pp. 113–130.

Nussbaum, Martha C. 2019. *The cosmopolitan tradition: A noble but flawed ideal*. Harvard University Press.

Oates, W.E. 1972. *Fiscal Federalism*. Harbrace series in business and economics Harcourt Brace Jovanovich.

Olken, Benjamin A. 2007. "Monitoring corruption: evidence from a field experiment in Indonesia." *Journal of political Economy* 115(2):200–249.

Patrinos, Harry Anthony, Felipe Barrera-Osorio and Tazeen Fasih. 2009. *Decentralized decision-making in schools: The theory and evidence on school-based management*. The World Bank.

Platteau, Jean-Philippe and Frederic Gaspart. 2003. "The risk of resource misappropriation in community-driven development." *World development* 31(10):1687–1703.

Pustejovsky, James E and Elizabeth Tipton. 2018. "Small-sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models." *Journal of Business & Economic Statistics* 36(4):672–683.

Raffler, Pia, Daniel N Posner and Doug Parkerson. 2019. "The weakness of bottom-up accountability: experimental evidence from the Ugandan health sector." *Technical Report, Harvard University and Innovations for Poverty Action* .

Rainey, Carlisle. 2014. "Arguing for a negligible effect." *American Journal of Political Science* 58(4):1083–1091.

Rashid, A. 2008. *Descent Into Chaos: The US and the Failure of Nation Building in Pakistan, Afghanistan, and Central Asia*. Viking.

Reinikka, Ritva and Jakob Svensson. 2004. "Local capture: evidence from a central government transfer program in Uganda." *The quarterly journal of economics* 119(2):679–705.

Rose, Pauline and Martin Greeley. 2006. *Education in fragile states: Capturing lessons and identifying good practice*. DAC Fragile States Group.

Rubenstein, Jennifer. 2015. *Between samaritans and states: The political ethics of humanitarian INGOs*. Oxford University Press.

Shleifer, Andrei and Robert W Vishny. 1993. "Corruption." *The quarterly journal of economics* 108(3):599–617.

Suhrke, Astri. 2011. *When More is Less: The International Project in Afghanistan*. Hurst.

Tanasoca, Ana and John S Dryzek. 2021. "Democratic altruism." *International Theory* 13(2):205–230.

Teele, Dawn Langan et al. 2014. "Reflections on the ethics of field experiments." *Field experiments and their critics: Essays on the uses and abuses of experimentation in the social sciences* pp. 115–140.

Wendler, David. 2021. The Ethics of Clinical Research. In *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. Winter 2021 ed. Metaphysics Research Lab, Stanford University.

Werker, Eric and Faisal Z. Ahmed. 2008. What do nongovernmental organizations do? In *J. Econ. Perspect.*

Wilson, James Q. 1961. "The economy of patronage." *Journal of political Economy* 69(4):369–380.

Wolfe, Rebecca. 2020. The benefits and challenges of randomized control trials in conflict environments: Reflections from a scholar-practitioner. In *Researching Peace, Conflict, and Power in the Field*. Springer pp. 259–280.

Appendix for "Can Communities Take Charge?"

A Additional Details on Geographic Context and Security Concerns

A.1 Kapisa and Parwan Provinces

Parwan and Kapisa lie north of Kabul. Approximately half of Kapisa's terrain is mountainous or semi-mountainous. According to the Central Statistical Organization (CSO) of Afghanistan, Kapisa has an estimated 441,000 residents, of which 99 percent live in rural areas where community-based schools are typically established. CARE implements CBE classes in four of Kapisa's seven districts, which are included in ALSE. These four districts (Mahmud Raqi, Hesa-e Dawm, Hesa-e Awli, and Kohband) are relatively more secure than the three (Tagab, Alasai, and Najrab). However, during the fall of 2016, ALSE was not able to carry out the capacity assessment surveys in eight villages in Kapisa due to security concerns.

More than two-thirds of Parwan province is mountainous or semi-mountainous, while the remaining area is flatlands. Parwan is slightly less rural than Kapisa and is estimated to have 664,500 inhabitants, of which about 75% live in rural areas and 25% in urban areas. CARE works in five of the ten districts in Parwan (Charikar, Syed Khel, Jabulsaraj, Salang, and Bagram) which are located in the central-east region of the province and are all in the ALSE sample. ALSE was able to conduct the data collection for the capacity assessment in all of the sampled villages in Parwan but one, which became inaccessible due to security concerns.

A.2 Central Highlands (Bamiyan, Daykundi, and Ghor Provinces)

Bamiyan is a mountain-locked province estimated to have 447,200 inhabitants, while the population of Daykundi and Ghor are 460,600 and 690,200 respectively, according to the CSO. ALSE conducted research in one or two districts in each of these provinces in the Central Highlands, where CRS has implemented CBE programs. These include the Yakawlang district in Bamiyan, Ashterlay and Sangi Takht districts in Daykundi; and the Lal wa Sarjangal district in Ghor. These districts are significantly less densely populated than those in Kapisa and Parwan and are spread over a much larger area. All three provinces are mostly mountainous or semi-mountainous and the vast majority of the populations live in rural areas, where of the three, Bamiyan has the greatest population living in urban areas (20%). We were able to administer the capacity assessment surveys in all villages sampled in these three provinces.

A.3 Herat Province

Herat is located in western Afghanistan and is one of the largest provinces with a population of about 1,780,000. One third of the province is mountainous or semi-mountainous and more than half is flatland. There are 16 districts in Herat Province, which is about 75% rural and

25% urban. CRS implements CBE programs in two districts in Herat (Adraskan and Gozara) and ALSE covers both of them. We were able to administer the capacity assessment surveys in all villages sampled in Herat.

Table A1: Summary Statistics on the Provinces and Districts Included in the ALSE

Province	**District**	**Geographic Area (square km)**	**Population Density (people per square km)**	**Number of Villages**
Kapisa	Mahmud Raqi	136	57.3	114
	Hesa-e-Dawn	237	188.3	91
	Hesa-e-Awli	1,908	30.8	98
	Kohband	149	157.4	195
Parwan	Charikar	191	933.7	117
	Syed Khel	47	1,352.50	136
	Jabulsarraj	485	166.3	99
	Salang	657	37.9	124
	Bagram	345	365.1	121
Bamiyan	Yakawlang	6,704	6.9	3.25
Daykundi	Ashterlay	1,360	34	320
	Sangi Takht	1,945	25.3	-
Ghor	Lal wa Sarjangal	5,115	32.6	676
Herat	Adraskan	10,070	13.9	384
	Gozara	2,000	200	400

Retrieved from Burde et al. (2017).

B Additional Details on Calculation of SESOI

The SESOI serves as a threshold beyond which we would declare that the sustainability model is inferior to an unacceptable degree in comparison to INGO administration of CBE. The derivation is as follows. Let μ_1 be the mean outcome under the experimental alternative (CBE), μ_0 be the mean outcome under the existing standard of care (INGO administration), μ_c be the mean outcome in the absence of any CBE, and then C_1 and C_0 are costs for the experimental alternative and standard of care, respectively. Then, we prefer the experimental alternative when it dominates in terms of cost-effectiveness:

$$\begin{aligned}
&\frac{\mu_1 - \mu_c}{C_1} > \frac{\mu_0 - \mu_c}{C_0} \\
&\Leftrightarrow \frac{\mu_1 - \mu_c}{\mu_0 - \mu_c} > \frac{C_1}{C_0} \\
&\Leftrightarrow \mu_1 > (\mu_0 - \mu_c)\frac{C_1}{C_0} + \mu_c \\
&\Leftrightarrow \mu_1 - \mu_0 > (\mu_0 - \mu_c)\frac{C_1}{C_0} - (\mu_0 - \mu_c) \\
&\Leftrightarrow \mu_1 - \mu_0 > -B \times (1 - C_1/C_0),
\end{aligned}$$

where $B = \mu_0 - \mu_c$, the treatment effect for the standard of care (INGO administration) relative to no CBE at all.

Our cost-effectiveness criterion is consistent with Hendren and Sprung-Keyser (2022)'s "marginal value of public funds" criterion under the assumptions that welfare gains are proportional to the measured effects ($\mu_1 - \mu_c$ and $\mu_0 - \mu_c$) and long-term effects on public costs are of negligible concern. Both of these assumptions are reasonable for the current case.

We used average treatment effect estimates from ALSE Phase One for B and a detailed cost comparison analysis (discussed below) for C_1 and C_0. The ATE estimates for school attendance and learning assessment test score outcomes are as follows: the Phase One ATE of INGO-administered CBE on the rate of school attendance is 0.139 (13.9 percentage points) and the Phase One ATE of INGO-administered CBE on learning assessment scores is 0.284 control group standard deviation (sd) units.

For other outcomes in our analysis, we did not have Phase One ATE estimates since these were not measured in Phase One. As a result, we used a rule of thumb that was based on what we observed for the school attendance and learning assessment scores. With regards to the learning assessment, we observed that the ATE was approximately 0.30 control group standard deviations. The ATE for school attendance, 0.139, was measured as a proportion. The Phase One control group rate of school attendance was 0.635 (63.5%) in which case an ATE of 0.139 is also about 0.30 control group standard deviations (note that if the attendance rate is 0.635, then, because attendance is a binary variable, the standard deviation equals 0.481). As such, we use the rule of thumb value of 0.30 control group standard deviations to establish our benchmark and set a threshold for outcomes other than school attendance and learning assessment. According to conventions in education research, 0.30 standard deviations would be considered a medium effect size.

C Cost-Comparison Analysis

During the ALSE Phase One study, we conducted a cost-effectiveness analysis (CEA) for NGO-administered CBE and presented the results in the ALSE Phase One Outcomes Report (Burde et al, 2016). For this Phase Two study, we compared costs by calculating the average CBE cost per village for three variations of CBE administration: NGO management, sustainability model - year 1, and sustainability model - projected year 2. Table A2 below shows all cost categories included in this calculation. Most expenses were originally reported in Afghanis. We used the average exchange rate (68.12 AFN per 1 USD) of the academic year (March 2018 to November 2018) for the "sustainability model-year 1" and "sustainability model-projected year 2" columns, whereas the NGO management costs were pulled in from our previous Phase One CEA.

We had collected NGO administration costs for years 1 and 2 of our Phase One study. Thus, for the purpose of this cost-comparison analysis, we used the average of these two years of NGO administration costs. As per the sustainability model-year 1, we identified the average CBE cost per village in the following manner: first, we estimated the total cost on a provincial level. These are the costs incurred throughout the handover year, when roughly half of the ALSE treatment villages were transferred from INGO administration to village-level community institutions. Then a total cost per village for each province was calculated relative to the number of handover villages in each province. The average cost per village in each of the six ALSE provinces was then calculated by multiplying the total cost per village by the provincial weight (number of handover villages in the province over the total number of handover villages in the ALSE sample). Ultimately, these averages were added to determine the average CBE cost per village. Sustainability model-projected year 2 refers to the second year following the CBE management transition. Estimates of the projected community management costs were based on the idea that these costs would look similar to the first year, but with a few exceptions, which are explained below. Seven cost categories were accounted for in this analysis:

1. Project staff: following the framework from Dhaliwal et al (2013), we included only national staff salaries. Contracts and payroll sheets were the key data sources for this category.

2. CBE class items: this category included classroom refurbishments, books for the classroom libraries, and winterization kits (these allow classrooms to prepare for winter temperatures).

3. Training: this category included all costs related to teaching training, village-level shura member training, and teacher learning groups. Teacher training costs were paid by the implementing INGOs. ALSE training costs for village shuras were accounted for in this category. We also include *mahram* (accompaniment) costs for female participants of the training workshops in this category, along with the teacher learning group cost (a capacity building activity that INGOS organized to connect CBE teachers with the nearest government school to create opportunities for exchange and learning).

4. Mobilization and monitoring visits: This included the cost of renting vehicles and paying for the fuel needed so mobilization and monitoring staff could visit the villages.

5. Supplies and materials: Classroom supplies and materials consisted of student kits, the teacher kit, classroom kits, and textbooks. The cost of these items was paid by the implementing INGOs.

6. Classroom rent: In most cases, the village community provided a physical space for use as a classroom, free of charge. However, in a few instances, INGOs rented the classroom for various reasons. This cost was included in this category too.

7. Teacher salaries: the information on teacher salaries was obtained from the implementing INGOs and ALSE project.

Table A2: Cost-Comparison Analysis

Cost Categories	Cost items	NGO Administration (Average of Year 1 and Year 2)		Sustainability Model (Year 1)		Sustainability Model (Projected Year 2)	
I. Project Staff	Direct and Indirect Staff	USD	2,641.67	USD	1,078.78	USD	1,078.78
II. Class items	Winterization Kits	USD	141.25	USD	-	USD	-
	Books for classroom libraries	USD	27.91	USD	-	USD	-
	Cabinet for the classroom libraries	USD	3.85	USD	-	USD	-
	Classroom refurbishments	USD	215.38	USD	11.44	USD	45.50
III. Trainings	**Teacher Training Costs**						
	Participation	USD	58.41			USD	41.44
	Transportation	USD	94.86			USD	19.01
	Meal	USD	182.96			USD	1.71
	Accomodation	USD	70.57	USD	41.44	USD	0.44
	Training Materials			USD	19.01	USD	1.89
				USD	1.71		
	Shura Trainings			USD	0.44		
	Participation	USD	94.81	USD	1.89	USD	53.01
	Transportation	USD	106.75			USD	115.36
	Meal	USD	37.67			USD	38.67
	Accommodation	USD	48.26	USD	159.04	USD	6.53
	Training Materials			USD	115.36	USD	2.37
				USD	116.02		
	Costs for Teacher Learning Groups	USD	17.36	USD	19.58		
	Transporation	USD	-	USD	2.37	USD	17.36
	Meal	USD	-				
	Accomodation	USD	14.91				
	Training Materials					USD	14.91
IV. Mobilization and Monitoring Visits	Rental Vehicle Costs						
	Fuel Costs	USD	1,143.11	USD	236.12	USD	236.12
V. Supplies and Materials	Student kits	USD	133.63	USD	179.17	USD	179.17
	Classroom kit	USD	75.49	USD	85.27	USD	85.27
	Math manipulative kits	USD	18.79	USD	-	USD	-
	Teacher kits	USD	19.37	USD	24.88	USD	24.88
	Sports kits	USD	4.71	USD	-	USD	-
	Textbooks	USD	94.64	USD	161.40	USD	161.40
VI. Classroom Rent		USD	255.88	USD	-	USD	-
VII. Teacher Salaries	Teacher salaries	USD	932.66	USD	1,202.85	USD	1,202.85
Average cost per village		**USD**	**6,438.88**	**USD**	**3,456.77**	**USD**	**3,326.68**

D Variable definitions and specifications

Table A3: Details for Outcomes Variables

Outcome	Data Source	Survey Questions
Children (Aged 8-13) Outcomes		
Attendance	Child Roster	Q22a, Q22b, Q22d, Q22e, Q25a, QQ25b, Q25d, Q25e
Math and Reading Score	Learning Assessment	Q5-10, Q1-4
Household Outcomes		
Perception on access to quality education	Household Survey	Q33a, Q33b, Q33c, Q33d
Confidence in local institutions	Household Survey	Q59a, Q59c, Q59d, Q59e
Teacher attendance	Household Survey	Q54
Community perception of teacher quality	Household Survey	Q55, Q56
Community Leader Outcomes		
Perceptions on access to quality education	Community Leader Survey	Q14a, Q14b, Q14c, Q14d
Classroom supplies	Community Leader Survey	Q20a, Q20b, Q48g, Q48h
Community provision for future classes	Community Leader Survey	Q30a-i==1, Q30a-i==3
Confidence in local community institutions	Community Leader Survey	Q31, Q32
Confidence that school will be sustained	Community Leader Survey	Q27, Q28, Q29
MoE provisions for funding teachers	Community Leader Survey	Q30a-i==5
Shura provisions for future classes	Community Leader Survey	Q30a-i==2
Teacher Outcomes		
Classroom supplies	Teacher Survey	Q34a, Q34b
Whether teacher is being paid regularly	Teacher Survey	Q16, Q17
Confidence in local community institutions	Teacher Survey	Q48a==2
Confidence that school will be sustained	Teacher Survey	Q47
Proactive management from MoE	Teacher Survey	Q32, Q33
Proactive management from shares	Teacher Survey	Q26, Q27, Q28, Q29, Q30, Q31
Teacher plans to remain CBE teacher	Teacher Survey	Q50, Q51
Follow-up Sustainability Outcomes		
Combined desire to continue CBE	Community Leader Follow-Up Survey	vA: Q8, vB: Q6
Initiative to have CBE sustained	Community Leader Follow-Up Survey	vA: Q6, vB: Q8
Provisions for sustainability without NGO	Community Leader Follow-Up Survey	vA: Q8, Q9, Q10; vB: Q6, Q12, Q13
Confidence that school will sustain	Community Leader Follow-Up Survey	vA: Q7, vB: Q11
Continued CBE operations in spring 2018	Community Leader Follow-Up Survey	Q1

Table A4: Covariate Specifications

Variable Name	Description	Cov1	Cov2
school_km	Distance to nearest formal school	X	X
ros_hhheadchild	Head of household's child (indicator)	X	X
ros_girl	Child is female (indicator)	X	X
ros_age	Child age	X	X
int_lang_pashto	Farsi/Pashto (indicator)	X	X
hh_eth_hazara	Ethnicity Hazara (indicator)	X	X
hh_eth_pashtun	Ethnicity Pashtun (indicator)	X	X
hh_eth_tajik	Ethnicity Tajik (indicator)	X	X
hhhead_age	Household head age	X	X
hhhead_schoolyears	Household head schooling (years)	X	X
hh_num	Number in household	X	X
hh_jeribs	Jeribs of land worked by household	X	X
hh_animal_sheep	Number of sheep	X	X
hhhead_job_farmer	Household head is a farmer (indicator)	X	X
hhhead_job_laborer	Household head is a laborer (indicator)		X
hhhead_job_military	Household head is in military (indicator)		X
hhhead_write	Household head can write (indicator)		X
hhhead_read	Household head can read (indicator)		X
hhhead_edu_mosque	Household head went to mosque school (indicator)	X	X
hhhead_edu_madrassa	Household head went to madrassa (indicator)		X
hhhead_edu_community	Household head to CBS (indicator)		X
hhhead_edu_government	Household head went to gov. school (indicator)		X
hhhead_edu_university	Household head went to university (indicator)		X
hhhead_female	Household head is female (indicator)		X
hh_totalincome_2000less	Total household income is 2000 AF or less		X
hh_totalincome_2001to5000	Total household income is 2001 to 5000 AF		X
hh_totalincome_5001to10000	Total household income is 5001 to 10000 AF		X
hh_totalincome_10001to15000	Total household income is 10001 to 15000 AF		X
hh_totalincome_15001plus	Total household income is 15001 AF or more		X
hhhead_shia	Household head is Shia		X
hhhead_sunni	Household head is Sunni		X
hh_own_tvs	Household has TV (indicator)		X
hh_own_mobiles	Household has mobile phone (indicator)		X
hh_own_cars	Household has car (indicator)		X
hh_own_radios	Household has radio (indicator)		X
factorwealth	A factor score measuring wealth		X
hh_landown	Household owns land (indicator)		X
hh_children	Number of children in household		X

E Descriptive tables

Table A5: Realized distribution of clusters, villages, household, and children over treatment and control conditions (after attrition)

	Sustainability Model Communities (Treatment Condition)	Extended NGO Administration Communities (Control Condition)
Clusters	53	51
Villages	64	68
Households	1778	1976
Children	2351	2629
Community Leaders	62	67
Teachers	70	84

Table A6: Covariate Balance Placebo Regressions

	Placebo Effect	SE	Unadj p-value
Child Covariates			
ros_hhheadchild	0.01	(0.02)	0.74
ros_girl	-0.02	(0.02)	0.29
ros_age	-0.04	(0.05)	0.44
Household Covariates			
int_lang_pashto	-0.01	(0.05)	0.81
hh_eth_hazara	-0.16	(0.11)	0.14
hh_eth_pashtun	0.00	(0.05)	0.96
hh_eth_tajik	0.14	(0.11)	0.21
hhhead_job_farmer	-0.03	(0.05)	0.52
hhhead_age	0.01	(0.04)	0.86
hhhead_schoolyears	0.14	(0.11)	0.24
hh_num	-0.02	(0.08)	0.77
hh_jeribs	0.05	(0.09)	0.59
hh_animal_sheep	-0.19	(0.19)	0.32
school_km	-0.33	(0.22)	0.14
hhhead_job_laborer	-0.03	(0.03)	0.34
hhhead_job_military	0.02	(0.02)	0.31
hhhead_write	0.02	(0.03)	0.46
hhhead_read	-0.01	(0.04)	0.71
hhhead_edu_mosque	0.02	(0.04)	0.69
hhhead_edu_madrassa	-0.04	(0.02)	0.09
hhhead_edu_community	-0.00	(0.01)	0.93
hhhead_edu_government	0.07	(0.05)	0.18
hhhead_edu_university	0.01	(0.01)	0.54
hhhead_female	-0.02	(0.01)	0.06
hh_totalincome_2000less	-0.04	(0.04)	0.34
hh_totalincome_2001to5000	-0.04	(0.03)	0.22
hh_totalincome_5001to10000	0.03	(0.03)	0.30
hh_totalincome_10001to15000	0.01	(0.02)	0.47
hh_totalincome_15001plus	0.01	(0.01)	0.43
hh_own_tvs	-0.09	(0.04)	0.04
hh_own_mobiles	-0.00	(0.02)	0.95
hh_own_cars	0.02	(0.03)	0.50
hh_own_radios	0.05	(0.04)	0.26
factorwealth	-0.05	(0.08)	0.54
hh_landown	-0.03	(0.04)	0.50
hh_children	-0.05	(0.07)	0.53
Province Indicators			
PROVINCE1	-0.01	(0.09)	0.93
PROVINCE2	0.15	(0.12)	0.21
PROVINCE3	-0.00	(0.05)	0.93
PROVINCE4	-0.02	(0.11)	0.87
PROVINCE5	-0.14	(0.08)	0.09
PROVINCE6	0.02	(0.08)	0.85

Table A7: Placebo Test of Differential Missingness Between Treatment and Control Conditions

	Treatment Effect	Std. Err.	p-value	N
Missing teacher attendance outcome	0.06	(0.04)	0.15	3754
Number of children tested in household	-0.02	(0.09)	0.83	3822
Number of children attending in household	-0.04	(0.12)	0.77	4942

F Checking for ethnic or gender-based bias or agency problems

Table A8: Heterogeneous Treatment Effects by Gender and Ethnic Status

Outcome	Group	Intercept	Treatment Effect	N	Het. test p-value
Attendance	Girls	0.76	-0.03	2183	0.88
		(0.02)	(0.03)		
	Boys	0.86	-0.01	2418	
		(0.02)	(0.03)		
Test Score	Girls	0.71	-0.06	1752	0.67
		(0.05)	(0.08)		
	Boys	0.85	0.01	1764	
		(0.04)	(0.06)		
Attendance	Ethnic majority	0.81	-0.03	4490	0.70
		(0.01)	(0.02)		
	Ethnic minority	0.78	0.01	490	
		(0.03)	(0.04)		
Test Score	Ethnic majority	0.79	-0.04	3430	0.97
		(0.03)	(0.06)		
	Ethnic minority	0.64	-0.08	392	
		(0.12)	(0.15)		
Attendance	CL same ethnicity	0.81	-0.02	4050	0.76
		(0.02)	(0.03)		
	CL diff ethnicity	0.78	0.01	715	
		(0.03)	(0.04)		
Test Score	CL same ethnicity	0.79	-0.04	3119	0.88
		(0.03)	(0.06)		
	CL diff ethnicity	0.56	0.04	553	
		(0.07)	(0.13)		

NOTE: Cluster-robust standard errors in parentheses. The heterogeneity test p-value is based on a Wald test of the null hypothesis that the group-level treatment effects are equal for each outcome. Variation in total sample sizes is due to variation in missingness in the group indicators.

Table A9: Heterogeneous Treatment Effects by Community Leader having School-age Children

Outcome	Group	Intercept	Treatment.Effect	N	Het.test.p.value
Attendance	Has children	0.81	-0.01	4000	0.72
		(0.01)	(0.02)		
	No children	0.85	-0.07	980	
		(0.06)	(0.07)		
Test score	Has children	0.79	-0.09	3082	0.54
		(0.04)	(0.07)		
	No children	0.66	0.12	740	
		(0.15)	(0.17)		
HH satisfaction with access	Has children	0.02	-0.03	3009	0.8
		(0.06)	(0.08)		
	No children	-0.05	0.08	741	
		(0.11)	(0.15)		
HH confidence in local instns.	Has children	0.02	0.03	3009	1
		(0.04)	(0.07)		
	No children	0.01	0.04	741	
		(0.09)	(0.12)		
HH perceived teacher attendance	Has children	0.01	0.02	2076	0.25
		(0.04)	(0.07)		
	No children	-0.45	0.52	465	
		(0.27)	(0.29)		
HH perceived teacher quality	Has children	0.03	-0.05	3009	0.98
		(0.04)	(0.06)		
	No children	-0.17	0.12	741	
		(0.14)	(0.16)		

G Checking for lingering capacity deficits

Table A10: Interaction Effects Regressions with Village-Level General Management Capacity Assessment [CA] Sum Scores with Main Child Outcomes and Agency Outcomes for Households, Leaders, and Teachers

Outcome	Treatment	CA sum score	Treatment×CA sum score	N
Child attendance	-0.03	0.02	-0.01	3894
	(0.04)	(0.02)	(0.02)	
Child test score	-0.03	0.09	-0.03	3003
	(0.09)	(0.04)	(0.06)	
HH confidence in local instns.	0.01	0.01	0.00	2932
	(0.08)	(0.06)	(0.07)	
HH perceived teacher quality	-0.14	0.01	0.00	2932
	(0.16)	(0.07)	(0.08)	
HH perceived teacher attendance	-0.09	0.16	-0.12	2027
	(0.11)	(0.07)	(0.08)	
CL classroom funds score	-0.27	0.19	-0.21	123
	(0.24)	(0.09)	(0.18)	
Teacher fund application index	0.28	-0.02	0.04	154
	(0.18)	(0.15)	(0.16)	
Teacher classroom supplies score	-0.04	0.05	-0.02	154
	(0.19)	(0.12)	(0.15)	
Teacher motivation index	-0.12	0.01	0.00	154
	(0.22)	(0.14)	(0.16)	
Teacher paid regularly	0.42	-0.08	0.07	154
	(0.21)	(0.16)	(0.17)	

NOTE: Cluster-robust standard errors in parentheses. The CA sum score is an index based on the sum of nine capacity assessment items measuring management capacity. The sum is standardized with respect to the pooled sample mean and standard deviation. All household and child results include the cov2 specification of controls. All leader and teacher outcomes include the cov1 specification of controls.

Table A11: Interaction Effects Regressions With Village-Level PCA Scores of Previous Capacity Building of CDCs with Main Child Outcomes and Agency Outcomes for Households, Leaders, and Teachers

Outcome	Treatment	CA PCA score	Treatment×CA PCA score	N
Child attendance	-0.04	-0.01	0.01	3894
	(0.04)	(0.01)	(0.01)	
Child test score	-0.03	0.01	-0.01	3003
	(0.09)	(0.02)	(0.02)	
HH confidence in local instns.	0.01	0.01	-0.01	2932
	(0.08)	(0.02)	(0.03)	
HH perceived teacher quality	-0.12	0.01	-0.05	2932
	(0.15)	(0.02)	(0.03)	
HH perceived teacher attendance	-0.09	0.05	-0.07	2027
	(0.1)	(0.03)	(0.04)	
CL classroom funds score	-0.3	0.08	-0.04	123
	(0.24)	(0.04)	(0.08)	
Teacher fund application index	0.3	-0.04	0.03	154
	(0.18)	(0.06)	(0.07)	
Teacher classroom supplies score	-0.01	-0.09	0.09	154
	(0.19)	(0.05)	(0.07)	
Teacher motivation index	-0.14	-0.06	0.13	154
	(0.22)	(0.06)	(0.07)	
Teacher paid regularly	0.43	0.02	-0.04	154
	(0.2)	(0.07)	(0.07)	

NOTE: Cluster-robust standard errors in parentheses. To construct the measure for the previous capacity building of CDCs, we use the first factor score from a principal component analysis (PCA) on ten questions on previous capacity building of CDCs. All household and child results include the cov2 specification of controls. All leader and teacher outcomes include the cov1 specification of controls.

Table A12: Interaction Effects Regressions With Village-Level PCA Scores of Current CDC Functionality with Main Child Outcomes and Agency Outcomes for Households, Leaders, and Teachers

Outcome	Treatment	CA PCA score	Treatment×CA PCA score	N
Child attendance	-0.04	-0.01	0.01	3894
	(0.04)	(0.01)	(0.01)	
Child test score	-0.02	0.03	-0.02	3003
	(0.1)	(0.02)	(0.03)	
HH confidence in local instns.	0.00	-0.01	-0.05	2932
	(0.08)	(0.03)	(0.04)	
HH perceived teacher quality	-0.15	-0.02	-0.05	2932
	(0.15)	(0.03)	(0.03)	
HH perceived teacher attendance	-0.11	-0.01	0.03	2027
	(0.11)	(0.03)	(0.05)	
CL classroom funds score	-0.29	0.05	-0.04	123
	(0.25)	(0.05)	(0.09)	
Teacher fund application index	0.31	-0.07	0.09	154
	(0.17)	(0.05)	(0.06)	
Teacher classroom supplies score	0.02	-0.12	0.2	154
	(0.19)	(0.05)	(0.08)	
Teacher motivation index	-0.08	-0.08	0.13	154
	(0.22)	(0.07)	(0.08)	
Teacher paid regularly	0.41	0.01	-0.05	154
	(0.20)	(0.07)	(0.07)	

NOTE: Cluster-robust standard errors in parentheses. To construct the measure for current CDC functionality, we use the first factor score from a principal components analysis on seven questions on CDC functionality. All household and child results include the cov2 specification of controls. All leader and teacher outcomes include the cov1 specification of controls.

H Effects on sustainability perceptions

Table A13: Effects of Sustainability Model on Sustainability Outcomes for Teachers and Community Leaders

Sample	Outcome	Treatment Effect	Std. Err.	p-value	N
Panel A: Community Leaders	Combined sustainability index	0.09	(0.17)	0.61	121
	Community provision for future classes	0.24	(0.18)	0.19	129
	Confidence in community institutions	-0.44	(0.2)	0.03	121
	Confidence that school will be sustained	0.08	(0.17)	0.64	121
	MOE provisions for funding teachers	-0.02	(0.19)	0.92	129
	Shura provisions for future classes	0.29	(0.19)	0.13	129
Panel B: Teachers	Combined sustainability index	-0.45	(0.21)	0.04	154
	Confidence in local community institutions	0.03	(0.04)	0.50	154
	Confidence that school will be sustained	-0.29	(0.08)	0.00	154
	Proactive management from MOE	-0.24	(0.18)	0.18	154
	Proactive management from shuras	-0.01	(0.19)	0.97	154
	Teacher plans to remain CBE teacher	-0.44	(0.19)	0.02	154

All results include the cov1 specification of controls. Cluster-robust standard errors in parentheses. p-values are unadjusted.

I Heterogeneity by INGO catchment area

In the main text, we present longer-term sustainability results for the subsample of communities (both treated and control) that were in the catchment area of the INGO (CARE) that *did not* extend programming after the trial period. Whether a village is in the catchment area of one or the other INGO is determined pre-treatment, and the decision of the one INGO to continue their programming was determined prior to the realization of our endline outcomes. So restricting the sample in this way does not undermine internal validity. But it does raise the concern that our findings from the subsample do not generalize to the villages that were excluded. To assess this, we examine heterogeneous treatment effects by implementing NGO for the main children's attendance and learning outcomes and household satisfaction outcome in Appendix Table A14. There are no substantial differences in effects across the communities serviced by two implementing NGOs, suggesting that the long-term sustainability analysis with only CARE communities are plausibly generalizable to the CRS communities.

Table A14: Heterogeneous Treatment Effects by Implementing NGO

Outcome	Treatment	CRS	Treatment×CRS	Interaction term p-value	N
Children's attendance	-0.01	0.05	-0.01	0.78	4980
	(0.03)	(0.07)	(0.05)		
Children's test scores	-0.03	0.08	0.03	0.81	3822
	(0.07)	(0.18)	(0.11)		
Household satisfaction	0.04	-0.3	-0.13	0.41	3750
	(0.06)	(0.2)	(0.16)		

Note: All results include the cov2 specification of controls. Cluster-robust standard errors in parentheses.

J Full sample results for long-term follow-up phone survey

The tables below include the villages that were in the catchment area of both INGOs, including the INGO (CRS) that extended their programming beyond the period the trial. By construction, the objective sustainability measure ("Continue CBE operation in spring 2018") scores higher for those villages, because access to schooling was guaranteed in ways that could not be guaranteed in the places served by either the sustainability model or the other INGO (CARE). This mechanical relationship is the reason for the large, negative treatment effect for the outcome in the last row, and also explains why the combined sustainability index effect is smaller than in the table presented in the main text.

Table A15: Community Leaders' Longer-term Perceptions of CBE Sustainability (CARE and CRS Samples)

	Treatment Effect	Std. Err.	N
Equivalency Outcome			
Community desire to continue CBE	-0.03	(0.02)	129
Non-Equivalency Outcomes			
Combined sustainability index	0.16	(0.20)	123
Initiative to have CBE sustained	0.59	(0.19)	129
Provisions for sustainability without NGO	0.28	(0.22)	123
Confidence that school will sustain	-0.13	(0.09)	129
Continued CBE operations in spring 2018	-0.33	(0.09)	129

Note: Cluster robust standard errors shown in parentheses. All regression use the "cov2" specification.

K How leaders perceptions differ by age and educational status

Table A16: Interaction Effect Regressions With Community Leaders' Age and Educational Status, Endline Community Leader Sustainability Perceptions

Outcome	Treatment	Older leader	Treatment×Older leader	N
Combined sustainability index	0.32	0.15	-0.45	121
	(0.24)	(0.32)	(0.31)	
Community provision for future classes	0.53	0.86	-0.61	129
	(0.26)	(0.41)	(0.38)	
Confidence in community institutions	0.14	-0.01	-1.09	121
	(0.25)	(0.32)	(0.34)	
Confidence that school will be sustained	0.19	-0.1	-0.2	121
	(0.26)	(0.34)	(0.38)	
MOE provisions for funding teachers	-0.18	-0.41	0.33	129
	(0.27)	(0.33)	(0.3)	
Shura provisions for future classes	0.13	0.5	0.28	129
	(0.26)	(0.34)	(0.34)	
Outcome	Treatment	Educated leader	Treatment×Educated leader	N
Combined sustainability index	0.09	0.12	-0.02	121
	(0.22)	(0.26)	(0.31)	
Community provision for future classes	0.34	0.08	-0.27	129
	(0.24)	(0.27)	(0.35)	
Confidence in community institutions	-0.49	-0.28	0.13	121
	(0.25)	(0.28)	(0.36)	
Confidence that school will be sustained	-0.02	0.11	0.23	121
	(0.24)	(0.25)	(0.36)	
MOE provisions for funding teachers	0.02	0.39	-0.11	129
	(0.23)	(0.27)	(0.34)	
Shura provisions for future classes	0.27	0.28	0.03	129
	(0.24)	(0.28)	(0.38)	

NOTE: "Older leader" is defined as having age above the median for community leaders (52 years old in our sample). "Educated" means having completed any formal education (39% of leaders in our sample). Cluster robust standard errors in parentheses. All regressions use the "cov1" covariate specification.

L Checking if leaders' pessimism about sustainability correlated with performance

Table A17: Interaction Effect Regressions With Community Mean Children's Test Score, Endline Community Leader Sustainability Perceptions

Outcome	Treatment	Community Mean Test Score	Treatment×Mean Test Score	N
Combined sustainability index	-0.14	-0.64	0.65	121
	(0.3)	(0.42)	(0.54)	
Community provision for future classes	0.11	-0.10	0.36	129
	(0.28)	(0.36)	(0.51)	
Confidence in community institutions	-0.76	-1.03	0.91	121
	(0.29)	(0.37)	(0.56)	
Confidence that school will be sustained	-0.05	-0.53	0.38	121
	(0.29)	(0.40)	(0.54)	
MOE provisions for funding teachers	0.06	0.02	-0.21	129
	(0.25)	(0.45)	(0.5)	
Shura provisions for future classes	0.06	0.12	0.65	129
	(0.25)	(0.45)	(0.55)	

NOTE: Cluster robust standard errors in parentheses. All regressions use the "cov1" covariate specification.

M Adherence to Ethical Research Principles

In carrying out this research, the research team has sought to adhere to the ethical research principles stated in (American Political Science Association, 2021*a*, Sec. H) and American Political Science Association (2021*b*). The New York University Institutional Review Board approved the study protocols in terms of protections of our human subjects' rights via ongoing informed and voluntary consent, non-coercion, risk mitigation, and confidentiality. Subjects were not compensated for their participation in the research. We employed a highly trained team of project associates to conduct randomized field audits to ensure that data collection personnel adhered to our protocols in the field.

Beyond standard human subjects protections, the study was designed in a manner that was highly sensitive to context-specific risks and vulnerabilities. We discuss ethical considerations for the design of the treatment conditions at length in the main text in the section explaining the randomized equivalency trial design. In the section on outcomes, we discuss our approach to assessing whether the intervention affected biases due to gender or ethnicity (our findings indicate that it did not).

The research was carried out in collaboration with agencies of the Government of the Islamic Republic of Afghanistan, non-governmental organizations, and international donor governments. In doing so, our work was regularly assessed to ensure compliance with all relevant laws and regulations within Afghanistan while also maintaining strong protections for individuals' rights. The funding from the US Agency for International Development and the Canadian International Development Agency was through grants with contracts that established the research team's intellectual freedom and unrestricted right to publish findings. The authors of this study declare that they maintain no conflicts of interest that would compromise the integrity of the research.

A replication archive for reproducing all results will be posted to a public database upon publication of the study.

Appendix References

American Political Science Association. 2021*a*. *A guide to professional ethics in political science*. American Political Science Association.

American Political Science Association. 2021*b*. *Principles and Guidance for Human Subjects Research*. American Political Science Association.

Hendren, Nathaniel and Ben Sprung-Keyser. 2022. "The Case for Using the MVPF in Empirical Welfare Analysis." *National Bureau of Economic Research* .

Supplementary Appendix for

"Can Communities Take Charge?"

This supplementary appendix includes the full regression output for the figures and regression table in the main body of the paper, as well as the regression tables in the main appendix.

A Full Results for Figure 2 in Main Paper

The following Table A1 include the full regression output that corresponds with the results presented in Figure 2 in the main paper. The first two columns correspond to the outcome of children's attendance, the middle two columns to the outcome of children's math and reading scores, and the last two columns to the outcome of household satisfaction. All regressions show cluster robust standard errors in parentheses.

Table A1: Primary Equivalence Results

	Attendance		Std Math and Reading Score		Household Satisfaction	
	Coef.	SE	Coef.	SE	Coef.	SE
Treatment	-0.02	(0.02)	-0.02	(0.05)	-0.01	(0.07)
Intercept	0.81	(0.01)	0.77	(0.03)	0.01	(0.05)
hhhead_child	0.00	(0.01)	0.06	(0.02)		
girl	-0.05	(0.01)	-0.09	(0.02)		
age	0.01	(0.01)	0.33	(0.02)		
int_lang_pashto	-0.02	(0.01)	-0.03	(0.04)	-0.09	(0.04)
hh_eth_hazara	0.03	(0.02)	0.14	(0.04)	-0.01	(0.04)
hh_eth_pashtun	-0.01	(0.02)	-0.02	(0.05)	0.06	(0.05)
hh_eth_tajik	0.01	(0.03)	0.10	(0.06)	0.11	(0.06)
hhhead_job_farmer	0.00	(0.01)	-0.04	(0.03)	0.02	(0.02)
hhhead_age	-0.01	(0.01)	0.02	(0.02)	0.00	(0.02)
hhhead_schoolyears	-0.01	(0.01)	0.00	(0.03)	0.02	(0.03)
hh_num	0.01	(0.01)	0.05	(0.03)	-0.04	(0.04)
hh_jeribs	0.01	(0.01)	0.01	(0.02)	-0.04	(0.02)
hh_animal_sheep	-0.01	(0.01)	-0.04	(0.02)	0.04	(0.02)
childrenschool_km	-0.02	(0.01)	-0.01	(0.04)		
hhhead_job_laborer	-0.02	(0.01)	-0.04	(0.02)	0.00	(0.03)
hhhead_job_military	0.00	(0.01)	-0.02	(0.02)	0.00	(0.02)
hhhead_write	0.01	(0.01)	0.03	(0.03)	0.00	(0.04)
hhhead_read	0.02	(0.01)	0.05	(0.03)	-0.02	(0.04)
hhhead_edu_mosque	0.02	(0.01)	0.04	(0.02)	-0.03	(0.03)
hhhead_edu_madrassa	-0.01	(0.01)	-0.01	(0.01)	0.00	(0.02)
hhhead_edu_community	0.00	(0.01)	0.01	(0.01)	0.03	(0.02)
hhhead_edu_government	0.00	(0.01)	0.07	(0.03)	0.02	(0.03)
hhhead_edu_university	0.00	(0.01)	0.00	(0.02)	-0.01	(0.03)
hhhead_female	-0.01	(0.01)	0.04	(0.02)	0.02	(0.02)
hh_totalincome_2001less	0.02	(0.02)	0.07	(0.05)	0.02	(0.07)
hh_totalincome_2001to5000	0.01	(0.02)	0.07	(0.04)	0.03	(0.06)
hh_totalincome_5001to10000	0.02	(0.02)	0.07	(0.04)	0.07	(0.06)
hh_totalincome_10001to15000	0.01	(0.02)	0.05	(0.03)	0.01	(0.04)
hh_totalincome_15001plus	0.01	(0.01)	0.04	(0.03)	0.04	(0.03)
hh_own_tvs	0.03	(0.01)	0.03	(0.02)	0.02	(0.02)
hh_own_mobiles	0.01	(0.01)	0.00	(0.02)	0.04	(0.02)
hh_own_cars	0.00	(0.01)	0.01	(0.02)	-0.01	(0.02)
hh_own_radios	0.01	(0.01)	-0.01	(0.02)	0.01	(0.02)
factor_wealth	-0.01	(0.02)	0.06	(0.04)	0.01	(0.03)
hh_landown	0.00	(0.01)	0.01	(0.02)	0.00	(0.02)
hh_children	-0.02	(0.01)	-0.05	(0.02)	0.10	(0.03)
PROVINCE1	0.03	(0.03)	-0.01	(0.07)	-0.06	(0.08)
PROVINCE2	-0.02	(0.03)	-0.05	(0.08)	0.19	(0.1)
PROVINCE3	0.00	(0.01)	0.00	(0.04)	-0.09	(0.04)
PROVINCE4	0.02	(0.01)	0.03	(0.03)	0.08	(0.06)
PROVINCE5	0.02	(0.01)	-0.02	(0.04)	0.02	(0.04)
hhschool_km					0.01	(0.04)
N	4980		3822		3750	

B Full Results for Figure 3 in Main Paper

The following Tables A6 - A5 include the full regression output that corresponds with the results presented in Figure 3 in the main paper. All tables show cluster robust standard errors in parentheses.

Table A2: Test of Potential Agency Issues: CL - Classroom Supplies

	CL: classroom supplies	
	Coef.	SE
Treatment	-0.09	(0.23)
Intercept	-0.01	(0.13)
cl_int_lang_pashto	-0.00	(0.09)
cl_hhhead_job_farmer	0.12	(0.11)
cl_hhhead_age	-0.16	(0.1)
cl_hhead_schoolyears	0.03	(0.11)
cl_school_km	0.15	(0.13)
N	121	

Table A3: Test of Potential Agency Issues: Results on Teacher Funds and Supplies

	TS: Combined Fund Application Index		TS: Classroom Supplies		TS: Teacher Paid Regularly	
	Coef.	SE	Coef.	SE	Coef.	SE
Treatment	0.42	(0.16)	0.07	(0.16)	0.52	(0.16)
Intercept	0.06	(0.14)	0.02	(0.11)	0.07	(0.14)
cv_ts	-0.26	(0.09)	-0.07	(0.07)	-0.30	(0.11)
N	154		154		154	

Table A4: Test of Potential Agency Issues - Results on Teacher Motivation

	TS: Combined Motivation Index		TS: Intrinsic Motivation		TS: Material Motivation and Satisfaction	
	Coef.	SE	Coef.	SE	Coef.	SE
Treatment	0.07	(0.17)	-0.16	(0.17)	0.20	(0.16)
Intercept	0.03	(0.13)	0.05	(0.13)	0.01	(0.12)
cv_ts	-0.14	(0.08)	-0.22	(0.08)	-0.06	(0.08)
N	154		154		154	

Table A5: Test of Potential Agency Issues: Results on HH Confidence and Perceptions of Teacher Quality and Attendance

	HH: Confidence in Local Institutions		HH: Teacher Attendance		HH: Perception of Teacher Quality	
	Coef.	SE	Coef.	SE	Coef.	SE
Treatment	0.04	(0.06)	0.09	(0.08)	-0.05	(0.09)
Intercept	0.02	(0.04)	-0.05	(0.06)	0.00	(0.06)
int_lang_pashto	-0.02	(0.04)	0.08	(0.03)	-0.03	(0.04)
hh_eth_hazara	0.05	(0.05)	0.10	(0.05)	0.06	(0.05)
hh_eth_pashtun	0.03	(0.06)	0.09	(0.06)	0.10	(0.05)
hh_eth_tajik	0.01	(0.08)	0.04	(0.09)	0.14	(0.07)
hhhead_job_farmer	0.03	(0.02)	-0.05	(0.03)	0.06	(0.02)
hhhead_age	0.02	(0.02)	0.05	(0.04)	0.00	(0.02)
hhhead_schoolyears	0.01	(0.03)	0.03	(0.04)	0.00	(0.04)
hh_num	0.05	(0.04)	-0.05	(0.03)	0.03	(0.04)
hh_jeribs	-0.04	(0.03)	0.01	(0.03)	-0.05	(0.02)
hh_animal_sheep	0.00	(0.02)	0.01	(0.03)	0.03	(0.03)
hhschool_km	-0.02	(0.04)	0.01	(0.04)	-0.01	(0.05)
hhhead_job_laborer	-0.02	(0.03)	-0.01	(0.03)	0.01	(0.02)
hhhead_job_military	-0.04	(0.02)	0.00	(0.03)	-0.04	(0.02)
hhhead_write	0.02	(0.04)	-0.06	(0.04)	-0.03	(0.04)
hhhead_read	-0.03	(0.04)	0.01	(0.04)	0.04	(0.04)
hhhead_edu_mosque	0.04	(0.03)	0.09	(0.03)	0.03	(0.03)
hhhead_edu_madrassa	0.01	(0.02)	-0.01	(0.03)	0.00	(0.02)
hhhead_edu_community	0.03	(0.02)	-0.01	(0.02)	0.01	(0.03)
hhhead_edu_government	0.02	(0.03)	0.02	(0.03)	0.00	(0.03)
hhhead_edu_university	0.02	(0.02)	0.05	(0.03)	0.02	(0.02)
hhhead_female	0.01	(0.02)	-0.05	(0.03)	0.01	(0.02)
hh_totalincome_2001less	0.13	(0.06)	-0.06	(0.05)	0.13	(0.06)
hh_totalincome_2001to5000	0.15	(0.06)	-0.02	(0.06)	0.14	(0.05)
hh_totalincome_5001to10000	0.18	(0.05)	-0.07	(0.05)	0.15	(0.05)
hh_totalincome_10001to15000	0.12	(0.03)	-0.03	(0.03)	0.07	(0.04)
hh_totalincome_15001plus	0.12	(0.03)	-0.07	(0.03)	0.08	(0.03)
hh_own_tvs	-0.01	(0.02)	-0.03	(0.02)	0.00	(0.02)
hh_own_mobiles	-0.01	(0.02)	0.05	(0.04)	0.01	(0.02)
hh_own_cars	-0.02	(0.03)	0.02	(0.02)	0.02	(0.03)
hh_own_radios	0.02	(0.02)	0.02	(0.02)	0.03	(0.02)
factor_wealth	0.04	(0.04)	-0.03	(0.07)	0.00	(0.04)
hh_landown	-0.02	(0.02)	-0.05	(0.03)	0.01	(0.02)
hh_children	-0.02	(0.03)	0.04	(0.04)	0.13	(0.03)
PROVINCE1	0.02	(0.09)	0.09	(0.08)	-0.19	(0.08)
PROVINCE2	0.23	(0.11)	0.02	(0.11)	0.14	(0.1)
PROVINCE3	-0.08	(0.05)	0.04	(0.04)	-0.10	(0.03)
PROVINCE4	0.02	(0.04)	0.07	(0.04)	0.04	(0.07)
PROVINCE5	0.02	(0.02)	0.13	(0.03)	0.01	(0.03)
N	3750		2541		3750	

C Full Results for Table 2 in Main Paper

The following Tables A6 and A7 include the full regression output that corresponds with the community-leader survey equivalency and non-equivalency outcomes presented in Table 2 in the main paper. Both tables show cluster robust standard errors in parentheses.

Table A6: Community Leaders' Longer-term Perceptions of CBE Sustainability: Equivalency Outcome

	Community Desire to Continue CBE	
	Coef.	SE
Treatment	-0.04	(0.04)
Intercept	1.00	(0.01)
CL_int_lang_pashto	0.01	(0.01)
CL_hhhead_job_farmer	-0.02	(0.02)
CL_hhead_age	-0.02	(0.02)
CL_clschool_km	-0.00	(0.01)
N	61	

Table A7: Community Leaders' Longer-Term Perceptions of CBE Sustainability - Non-Equivalency Outcomes

	Combined Sustainability Index		Initiative to have CBE Sustained		Provisions for Sustain. Without NGO		Confidence that School will Sustain		Continued CBE Opts. in Spring 2018	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Treatment	0.44	(0.31)	0.48	(0.31)			0.05	(0.13)	-0.03	(0.12)
Intercept	-0.21	(0.21)	0.12	(0.23)	0.03	(0.25)	0.45	(0.09)	0.32	(0.08)
CL_int_lang_pashto	0.21	(0.19)	0.19	(0.17)	0.16	(0.23)	0.05	(0.08)	0.01	(0.08)
CL_hhhead_job_farmer	0.07	(0.18)	0.14	(0.2)	0.09	(0.21)	-0.02	(0.08)	-0.03	(0.08)
CL_hhead_age	-0.03	(0.16)	-0.37	(0.2)	0.21	(0.18)	0.05	(0.07)	0.04	(0.08)
CL_clschool_km	0.54	(0.12)	0.11	(0.12)	0.56	(0.22)	0.13	(0.05)	0.20	(0.04)
Treatment					0.30	(0.34)				
N	60		61		60		61		61	

D Full Results for Table A8 in Main Appendix

The following Tables A8-A10 include the full regression output that corresponds with Table A8 in the main appendix on heterogeneous treatment effects by gender, ethnic majority/minority status, and sharing the same vs. diff ethnicity as the community leader for children's attendance and test scores. For each of these tables, please note that cluster-robust standard errors are in parentheses. The heterogeneity test p-value is based on a Walt test of the null hypothesis that the group-level treatment effects are equal for each outcome.

Table A8: Heterogeneous Treatment Effects for Attendance and Test Scores by Gender

	Attendance			Math and Reading Scores		
	Coef.	SE	Het. test p-value	Coef.	SE	Het. test p-value
Girls - Intercept	0.76	(0.02)	0.88	0.71	(0.05)	0.67
Girls - Treatment	-0.03	(0.03)		-0.06	(0.08)	
Boys - Intercept	0.86	(0.02)		0.85	(0.04)	
Boys - Treatment	-0.01	(0.03)		0.01	(0.06)	
int_lang_pashto	-0.04	(0.01)		-0.12	(0.03)	
hhhead_job_farmer	-0.01	(0.01)		-0.04	(0.03)	
hhhead_age	-0.01	(0.01)		0.04	(0.02)	
hhhead_schoolyears	0.00	(0.01)		0.05	(0.03)	
hh_num	0.01	(0.01)		0.05	(0.03)	
hh_jeribs	0.01	(0.01)		0.01	(0.02)	
hh_animal_sheep	-0.01	(0.01)		-0.06	(0.02)	
school_km	-0.02	(0.01)		0.02	(0.04)	
hhhead_job_laborer	-0.02	(0.01)		-0.04	(0.02)	
hhhead_job_military	0.00	(0.01)		-0.02	(0.02)	
hhhead_write	0.00	(0.01)		0.03	(0.04)	
hhhead_read	0.03	(0.01)		0.04	(0.04)	
hhhead_edu_mosque	0.02	(0.01)		0.03	(0.03)	
hhhead_edu_madrassa	0.00	(0.01)		-0.01	(0.02)	
hhhead_edu_community	-0.01	(0.01)		0.01	(0.01)	
hhhead_edu_government	0.00	(0.01)		0.08	(0.04)	
hhhead_edu_university	0.00	(0.01)		0.01	(0.02)	
hhhead_female	-0.01	(0.01)		0.04	(0.02)	
hh_totalincome_2000less	0.02	(0.02)		0.06	(0.05)	
hh_totalincome_2001to5000	0.01	(0.02)		0.06	(0.04)	
hh_totalincome_5001to10000	0.01	(0.02)		0.06	(0.05)	
hh_totalincome_10001to15000	0.00	(0.02)		0.05	(0.03)	
hh_totalincome_15001plus	0.01	(0.01)		0.02	(0.03)	
hh_own_tvs	0.03	(0.01)		0.03	(0.02)	
hh_own_mobiles	0.00	(0.01)		0.00	(0.02)	
hh_own_cars	0.00	(0.01)		0.00	(0.02)	
hh_own_radios	0.01	(0.01)		-0.02	(0.02)	
factorwealth	-0.01	(0.02)		0.07	(0.04)	
hh_landown	0.00	(0.01)		0.01	(0.03)	
hh_children	-0.02	(0.01)		-0.08	(0.03)	
PROVINCE1	0.01	(0.02)		-0.05	(0.05)	
PROVINCE2	-0.04	(0.02)		-0.07	(0.05)	
PROVINCE3	-0.01	(0.01)		-0.04	(0.04)	
PROVINCE4	0.02	(0.01)		0.03	(0.04)	
PROVINCE5	0.02	(0.01)		-0.02	(0.04)	
N	4601			3516		

Table A9: Heterogeneous Treatment Effects for Attendance and Test Scores by Ethnic Majority/Minority Status

	Attendance			Math and Reading Scores		
	Coef.	SE	Het. test p-value	Coef.	SE	Het. test p-value
Ethnic Majority - Intercept	0.81	(0.01)	0.70	0.79	(0.03)	0.97
Ethnic Majority - Treatment	-0.03	(0.02)		-0.04	(0.06)	
Ethnic Minority - Intercept	0.78	(0.03)		0.64	(0.12)	
Ethnic Minority - Treatment	0.01	(0.04)		-0.08	(0.15)	
int_lang_pashto	-0.03	(0.01)		-0.12	(0.03)	
hhhead_job_farmer	0.00	(0.01)		-0.03	(0.03)	
hhhead_age	-0.01	(0.01)		0.03	(0.02)	
hhhead_schoolyears	-0.01	(0.01)		0.03	(0.03)	
hh_num	0.01	(0.01)		0.05	(0.03)	
hh_jeribs	0.01	(0.01)		0.01	(0.02)	
hh_animal_sheep	-0.01	(0.01)		-0.03	(0.02)	
school_km	-0.02	(0.01)		-0.01	(0.05)	
hhhead_job_laborer	-0.02	(0.01)		-0.04	(0.02)	
hhhead_job_military	0.00	(0.01)		-0.03	(0.02)	
hhhead_write	0.01	(0.01)		0.04	(0.04)	
hhhead_read	0.02	(0.01)		0.03	(0.04)	
hhhead_edu_mosque	0.02	(0.01)		0.03	(0.02)	
hhhead_edu_madrassa	-0.01	(0.01)		-0.01	(0.02)	
hhhead_edu_community	-0.01	(0.01)		0.02	(0.01)	
hhhead_edu_government	0.01	(0.01)		0.08	(0.04)	
hhhead_edu_university	0.00	(0.01)		0.01	(0.02)	
hhhead_female	-0.01	(0.01)		0.03	(0.02)	
hh_totalincome_2000less	0.02	(0.02)		0.06	(0.05)	
hh_totalincome_2001to5000	0.01	(0.02)		0.06	(0.04)	
hh_totalincome_5001to10000	0.02	(0.02)		0.06	(0.04)	
hh_totalincome_10001to15000	0.01	(0.01)		0.06	(0.03)	
hh_totalincome_15001plus	0.01	(0.01)		0.03	(0.03)	
hh_own_tvs	0.03	(0.01)		0.03	(0.02)	
hh_own_mobiles	0.01	(0.01)		0.01	(0.02)	
hh_own_cars	0.00	(0.01)		0.01	(0.02)	
hh_own_radios	0.01	(0.01)		-0.01	(0.02)	
factorwealth	-0.01	(0.02)		0.07	(0.04)	
hh_landown	0.00	(0.01)		0.02	(0.03)	
hh_children	-0.02	(0.01)		-0.08	(0.02)	
PROVINCE1	0.01	(0.02)		-0.05	(0.05)	
PROVINCE2	-0.04	(0.02)		-0.08	(0.05)	
PROVINCE3	-0.01	(0.01)		-0.03	(0.04)	
PROVINCE4	0.02	(0.01)		0.02	(0.04)	
PROVINCE5	0.01	(0.01)		-0.03	(0.04)	
N	4980			3822		

Table A10: Heterogeneous Treatment Effects for Attendance and Test Scores by Community Leader Ethnic Status

	Attendance			Math and Reading Scores		
	Coef.	SE	Het. test p-value	Coef.	SE	Het. test p-value
CL Same Ethnicity - Intercept	0.81	(0.02)	0.76	0.79	(0.03)	0.88
CL Same Ethnicity - Treatment	-0.02	(0.03)		-0.04	(0.06)	
CL Diff Ethnicity - Intercept	0.78	(0.03)		0.56	(0.07)	
CL Diff Ethnicity - Treatment	0.01	(0.04)		0.04	(0.13)	
int_lang_pashto	-0.03	(0.01)		-0.11	(0.04)	
hhhead_job_farmer	-0.01	(0.01)		-0.05	(0.03)	
hhhead_age	-0.01	(0.01)		0.03	(0.02)	
hhhead_schoolyears	0.00	(0.01)		0.03	(0.03)	
hh_num	0.01	(0.01)		0.06	(0.03)	
hh_jeribs	0.01	(0.01)		0.02	(0.02)	
hh_animal_sheep	-0.01	(0.01)		-0.03	(0.02)	
school_km	-0.02	(0.01)		0.00	(0.05)	
hhhead_job_laborer	-0.02	(0.01)		-0.05	(0.02)	
hhhead_job_military	0.00	(0.01)		-0.03	(0.02)	
hhhead_write	0.01	(0.01)		0.03	(0.04)	
hhhead_read	0.02	(0.01)		0.03	(0.04)	
hhhead_edu_mosque	0.02	(0.01)		0.03	(0.02)	
hhhead_edu_madrassa	-0.01	(0.01)		-0.01	(0.02)	
hhhead_edu_community	-0.01	(0.01)		0.01	(0.01)	
hhhead_edu_government	0.01	(0.01)		0.09	(0.04)	
hhhead_edu_university	0.00	(0.01)		0.01	(0.02)	
hhhead_female	-0.01	(0.01)		0.03	(0.02)	
hh_totalincome_2000less	0.01	(0.02)		0.04	(0.05)	
hh_totalincome_2001to5000	0.00	(0.02)		0.04	(0.05)	
hh_totalincome_5001to10000	0.01	(0.02)		0.04	(0.05)	
hh_totalincome_10001to15000	0.00	(0.02)		0.04	(0.03)	
hh_totalincome_15001plus	0.00	(0.01)		0.01	(0.03)	
hh_own_tvs	0.03	(0.01)		0.03	(0.02)	
hh_own_mobiles	0.01	(0.01)		0.00	(0.02)	
hh_own_cars	0.00	(0.01)		0.00	(0.02)	
hh_own_radios	0.01	(0.01)		-0.02	(0.02)	
factorwealth	-0.01	(0.02)		0.07	(0.04)	
hh_landown	0.00	(0.01)		0.03	(0.03)	
hh_children	-0.02	(0.01)		-0.08	(0.02)	
PROVINCE1	0.01	(0.02)		-0.04	(0.05)	
PROVINCE2	-0.04	(0.02)		-0.08	(0.05)	
PROVINCE3	-0.02	(0.01)		-0.07	(0.02)	
PROVINCE4	0.02	(0.01)		0.02	(0.04)	
PROVINCE5	0.01	(0.01)		-0.03	(0.04)	
N	3822			4765		

E Full Results for Table A9 in Main Appendix

The following Tables A11 and A12 include the full regression output that corresponds with Table A9 in the main appendix on heterogeneous treatment effects by the community leader having school-age children, for children's attendance and test scores, as well as household satisfaction with access, confidence in local institution, perceived teacher attendance, and perceived teacher quality. For each of these tables, please note that cluster-robust standard errors are in parentheses. The heterogeneity test p-value is based on a Walt test of the null hypothesis that the group-level treatment effects are equal for each outcome.

Table A11: Heterogeneous Treatment Effects for Attendance and Test Scores by Community Leader Having School-Aged Children

	Attendance			Math and Reading Scores		
	Coef.	SE	Het. test p-value	Coef.	SE	Het. test p-value
Has Children - Intercept	0.81	(0.01)	0.72	0.79	(0.04)	0.54
Has Children - Treatment	-0.01	(0.02)		-0.09	(0.07)	
No Children - Intercept	0.85	(0.06)		0.66	(0.15)	
No Children - Treatment	-0.07	(0.07)		0.12	(0.17)	
int_lang_pashto	-0.03	(0.01)		-0.11	(0.04)	
hhhead_job_farmer	0.00	(0.01)		-0.04	(0.03)	
hhhead_age	-0.01	(0.01)		0.03	(0.02)	
hhhead_schoolyears	-0.01	(0.01)		0.03	(0.03)	
hh_num	0.01	(0.01)		0.05	(0.03)	
hh_jeribs	0.01	(0.01)		0.01	(0.02)	
hh_animal_sheep	-0.01	(0.01)		-0.03	(0.02)	
school_km	-0.02	(0.01)		-0.01	(0.06)	
hhhead_job_laborer	-0.02	(0.01)		-0.05	(0.02)	
hhhead_job_military	0.00	(0.01)		-0.02	(0.02)	
hhhead_write	0.01	(0.01)		0.04	(0.04)	
hhhead_read	0.02	(0.01)		0.02	(0.04)	
hhhead_edu_mosque	0.02	(0.01)		0.03	(0.02)	
hhhead_edu_madrassa	-0.01	(0.01)		-0.01	(0.02)	
hhhead_edu_community	-0.01	(0.01)		0.02	(0.01)	
hhhead_edu_government	0.01	(0.01)		0.08	(0.04)	
hhhead_edu_university	0.00	(0.01)		0.01	(0.02)	
hhhead_female	-0.01	(0.01)		0.04	(0.02)	
hh_totalincome_2000less	0.02	(0.02)		0.07	(0.05)	
hh_totalincome_2001to5000	0.01	(0.02)		0.06	(0.04)	
hh_totalincome_5001to10000	0.02	(0.02)		0.07	(0.05)	
hh_totalincome_10001to15000	0.00	(0.02)		0.06	(0.03)	
hh_totalincome_15001plus	0.01	(0.01)		0.03	(0.03)	
hh_own_tvs	0.03	(0.01)		0.03	(0.02)	
hh_own_mobiles	0.01	(0.01)		0.00	(0.02)	
hh_own_cars	0.00	(0.01)		0.01	(0.02)	
hh_own_radios	0.01	(0.01)		-0.01	(0.02)	
factorwealth	-0.01	(0.02)		0.07	(0.04)	
hh_landown	0.00	(0.01)		0.02	(0.03)	
hh_children	-0.02	(0.01)		-0.08	(0.02)	
PROVINCE1	0.01	(0.02)		-0.06	(0.05)	
PROVINCE2	-0.04	(0.02)		-0.09	(0.06)	
PROVINCE3	-0.01	(0.01)		-0.03	(0.04)	
PROVINCE4	0.02	(0.01)		0.02	(0.04)	
PROVINCE5	0.02	(0.01)		-0.03	(0.04)	
N	4980			3822		

Table A12: Heterogeneous Treatment Effects for Household Outcomes by Community Leader Having School-Aged Children

	HH Satisfaction With Access			HH Confidence in Local Instns.			HH Perceived Teacher Attendance			HH Perceived Teacher Quality		
	Coef.	SE	Het. test p-value	Coef.	SE	Het. test p-value	Coef.	SE	Het. test p-value	Coef.	SE	Het. test p-value
Has Children - Intercept	0.02	(0.06)	0.8	0.02	(0.04)	1	0.01	(0.04)	0.25	0.03	(0.04)	0.98
Has Children - Treatment	-0.03	(0.08)		0.03	(0.07)		0.02	(0.07)		-0.05	(0.06)	
No Children - Intercept	-0.05	(0.11)		0.01	(0.09)		-0.45	(0.27)		-0.17	(0.14)	
No Children - Treatment	0.08	(0.15)		0.04	(0.12)		0.52	(0.29)		0.12	(0.16)	
int_lang_pashto	-0.10	(0.02)		0.00	(0.02)		0.13	(0.03)		-0.02	(0.02)	
hhhead_job_farmer	0.02	(0.02)		0.04	(0.02)		-0.05	(0.03)		0.04	(0.01)	
hhhead_age	0.00	(0.02)		0.02	(0.02)		0.05	(0.04)		0.00	(0.01)	
hhhead_schoolyears	0.02	(0.03)		0.01	(0.03)		0.02	(0.04)		0.00	(0.02)	
hh_num	-0.05	(0.04)		0.05	(0.04)		-0.06	(0.03)		0.02	(0.02)	
hh_jeribs	-0.04	(0.02)		-0.04	(0.03)		0.01	(0.03)		-0.03	(0.01)	
hh_animal_sheep	0.04	(0.03)		0.00	(0.02)		0.03	(0.03)		0.02	(0.02)	
school_km	0.01	(0.04)		-0.02	(0.04)		0.01	(0.04)		0.00	(0.03)	
hhhead_job_laborer	0.00	(0.03)		-0.02	(0.03)		-0.01	(0.03)		0.01	(0.01)	
hhhead_job_military	0.00	(0.02)		-0.04	(0.02)		0.00	(0.03)		-0.02	(0.01)	
hhhead_write	0.00	(0.04)		0.02	(0.04)		-0.06	(0.04)		-0.01	(0.02)	
hhhead_read	-0.02	(0.04)		-0.03	(0.04)		0.01	(0.04)		0.02	(0.03)	
hhhead_edu_mosque	-0.03	(0.03)		0.04	(0.03)		0.09	(0.03)		0.02	(0.02)	
hhhead_edu_madrassa	0.00	(0.02)		0.01	(0.02)		-0.01	(0.03)		0.00	(0.01)	
hhhead_edu_community	0.03	(0.02)		0.03	(0.02)		-0.01	(0.02)		0.01	(0.01)	
hhhead_edu_government	0.02	(0.03)		0.02	(0.03)		0.02	(0.03)		0.00	(0.02)	
hhhead_edu_university	-0.01	(0.03)		0.02	(0.02)		0.05	(0.02)		0.01	(0.01)	
hhhead_female	0.02	(0.02)		0.01	(0.02)		-0.05	(0.03)		0.00	(0.01)	
hh_totalincome_2000less	0.01	(0.07)		0.13	(0.06)		-0.06	(0.05)		0.08	(0.04)	
hh_totalincome_2001to5000	0.03	(0.06)		0.16	(0.06)		-0.02	(0.06)		0.09	(0.03)	
hh_totalincome_5001to10000	0.07	(0.06)		0.18	(0.05)		-0.07	(0.05)		0.10	(0.03)	
hh_totalincome_10001to15000	0.01	(0.04)		0.12	(0.03)		-0.02	(0.04)		0.05	(0.02)	
hh_totalincome_15001plus	0.04	(0.03)		0.12	(0.03)		-0.06	(0.04)		0.05	(0.02)	
hh_own_tvs	0.02	(0.02)		-0.01	(0.02)		-0.03	(0.02)		0.00	(0.02)	
hh_own_mobiles	0.04	(0.02)		-0.01	(0.03)		0.04	(0.04)		0.00	(0.01)	
hh_own_cars	-0.01	(0.02)		-0.02	(0.02)		0.02	(0.02)		0.01	(0.01)	
hh_own_radios	0.01	(0.02)		0.02	(0.02)		0.02	(0.02)		0.02	(0.01)	
factorwealth	0.01	(0.03)		0.04	(0.04)		-0.03	(0.07)		0.00	(0.02)	
hh_landown	0.00	(0.02)		-0.02	(0.02)		-0.05	(0.02)		0.01	(0.01)	
hh_children	0.10	(0.03)		-0.02	(0.03)		0.04	(0.04)		0.08	(0.02)	
PROVINCE1	0.03	(0.05)		-0.01	(0.04)		0.04	(0.05)		-0.08	(0.03)	
PROVINCE2	0.29	(0.06)		0.20	(0.05)		-0.02	(0.07)		0.14	(0.04)	
PROVINCE3	-0.05	(0.02)		-0.09	(0.03)		0.04	(0.03)		-0.05	(0.01)	
PROVINCE4	0.09	(0.06)		0.02	(0.04)		0.07	(0.05)		0.02	(0.04)	
PROVINCE5	0.02	(0.04)		0.02	(0.02)		0.14	(0.04)		0.00	(0.02)	
N	3750			3750			2541			3750		

F Full Results for Table A10 in Main Appendix

The following Tables A13-A16 include the full regression output that corresponds with Table A10 in the main appendix on interaction effects regressions with village-level general management capacity assessment sum scores with main child outcomes and agency outcomes for households, leaders, and teachers. For each of these tables, please note that cluster-robust standard errors are in parentheses. The CA sum score is an index based on the sum of nine capacity assessment items measuring management capacity. The sum is standardized with respect to the pooled sample mean and standard deviation

Table A13: Interaction Effects Regressions with Village-Level General Management Capacity Assessment [CA] Sum Scores with Main Child Outcomes

	Attendance		Child Test Score	
	Coef.	SE	Coef.	SE
Treatment	-0.03	(0.04)	-0.03	(0.09)
CA sum score	0.02	(0.02)	0.09	(0.04)
Treatment X CA sum score	-0.01	(0.02)	-0.03	(0.06)
Intercept	0.82	(0.02)	0.74	(0.04)
HAND_kapisa	-0.03	(0.07)	0.07	(0.16)
HAND_parwan	0.11	(0.06)	0.13	(0.15)
ros_hhheadchild	0.00	(0.01)	0.05	(0.02)
ros_girl	-0.05	(0.01)	-0.08	(0.02)
ros_age	0.01	(0.01)	0.32	(0.02)
int_lang_pashto	-0.02	(0.02)	-0.01	(0.06)
hh_eth_hazara	0.02	(0.02)	0.13	(0.04)
hh_eth_pashtun	-0.05	(0.03)	-0.07	(0.05)
hh_eth_tajik	-0.02	(0.03)	0.09	(0.06)
hhhead_job_farmer	0.00	(0.01)	-0.03	(0.03)
hhhead_age	-0.01	(0.01)	0.03	(0.02)
hhhead_schoolyears	0.00	(0.02)	-0.02	(0.03)
hh_num	0.01	(0.02)	0.08	(0.03)
hh_jeribs	0.00	(0.01)	0.01	(0.02)
hh_animal_sheep	-0.01	(0.01)	-0.06	(0.03)
school_km	-0.03	(0.02)	-0.04	(0.05)
hhhead_job_laborer	-0.02	(0.01)	-0.05	(0.03)
hhhead_job_military	0.00	(0.01)	-0.03	(0.02)
hhhead_write	0.01	(0.01)	0.03	(0.03)
hhhead_read	0.02	(0.01)	0.03	(0.04)
hhhead_edu_mosque	0.02	(0.01)	0.07	(0.02)
hhhead_edu_madrassa	0.00	(0.01)	-0.03	(0.02)
hhhead_edu_community	-0.01	(0.01)	0.01	(0.01)
hhhead_edu_government	0.00	(0.01)	0.10	(0.04)
hhhead_edu_university	0.00	(0.01)	0.00	(0.02)
hhhead_female	-0.02	(0.01)	0.04	(0.02)
hh_totalincome_2000less	0.01	(0.02)	0.10	(0.06)
hh_totalincome_2001to5000	0.00	(0.02)	0.09	(0.06)
hh_totalincome_5001to10000	0.03	(0.02)	0.12	(0.06)
hh_totalincome_10001to15000	0.00	(0.02)	0.07	(0.04)
hh_totalincome_15001plus	0.00	(0.01)	0.07	(0.04)
hh_own_tvs	0.02	(0.01)	0.02	(0.02)
hh_own_mobiles	0.00	(0.01)	0.00	(0.02)
hh_own_cars	-0.01	(0.01)	0.00	(0.02)
hh_own_radios	0.00	(0.01)	-0.01	(0.02)
factorwealth	0.00	(0.02)	0.06	(0.04)
hh_landown	0.01	(0.01)	0.02	(0.02)
hh_children	-0.02	(0.01)	-0.06	(0.03)
PROVINCE1	0.01	(0.04)	-0.03	(0.07)
PROVINCE2	0.01	(0.05)	-0.06	(0.1)
PROVINCE3	0.02	(0.01)	0.02	(0.04)
PROVINCE4	0.02	(0.02)	0.05	(0.04)
PROVINCE5	0.02	(0.01)	-0.01	(0.04)
N	3894		3003	

Table A14: Interaction Effects Regressions with Village-Level General Management Capacity Assessment [CA] Sum Scores with Agency Outcomes for Households

	HH Conf. in Local Instns.		HH Perceived Teacher Quality		HH Perceived Teacher Attendance	
	Coef.	SE	Coef.	SE	Coef.	SE
Treatment	0.01	(0.08)	-0.14	(0.16)	-0.09	(0.11)
CA sum score	0.01	(0.06)	0.01	(0.07)	0.16	(0.07)
Treatment X CA sum score	0.00	(0.07)	0.00	(0.08)	-0.12	(0.08)
Intercept	0.03	(0.04)	-0.05	(0.07)	-0.07	(0.09)
HAND_kapisa	0.01	(0.13)	0.56	(0.24)	0.48	(0.31)
HAND_parwan	0.14	(0.18)	-0.03	(0.2)	0.22	(0.17)
int_lang_pashto	-0.01	(0.04)	0.08	(0.04)	0.13	(0.07)
hh_eth_hazara	0.04	(0.04)	0.03	(0.05)	0.06	(0.06)
hh_eth_pashtun	0.04	(0.06)	0.11	(0.06)	0.07	(0.06)
hh_eth_tajik	0.03	(0.08)	0.16	(0.08)	0.04	(0.09)
hhhead_job_farmer	0.03	(0.02)	0.05	(0.02)	-0.03	(0.04)
hhhead_age	0.02	(0.02)	0.01	(0.02)	0.04	(0.04)
hhhead_schoolyears	0.02	(0.04)	0.00	(0.04)	0.03	(0.05)
hh_num	0.06	(0.04)	0.04	(0.04)	-0.02	(0.04)
hh_jeribs	-0.06	(0.03)	-0.06	(0.03)	0.04	(0.03)
hh_animal_sheep	0.00	(0.03)	0.03	(0.03)	0.01	(0.04)
school_km	-0.04	(0.03)	-0.04	(0.05)	-0.03	(0.05)
hhhead_job_laborer	-0.03	(0.03)	0.02	(0.03)	-0.01	(0.03)
hhhead_job_military	-0.05	(0.03)	-0.03	(0.02)	-0.01	(0.04)
hhhead_write	0.00	(0.05)	-0.01	(0.04)	-0.09	(0.04)
hhhead_read	-0.03	(0.05)	0.04	(0.04)	0.03	(0.04)
hhhead_edu_mosque	0.06	(0.04)	0.04	(0.03)	0.07	(0.03)
hhhead_edu_madrassa	0.03	(0.03)	-0.01	(0.03)	0.00	(0.04)
hhhead_edu_community	0.03	(0.02)	0.01	(0.03)	0.00	(0.02)
hhhead_edu_government	0.00	(0.04)	-0.02	(0.04)	0.04	(0.04)
hhhead_edu_university	0.03	(0.02)	0.03	(0.02)	0.03	(0.03)
hhhead_female	0.01	(0.02)	0.00	(0.02)	-0.06	(0.03)
hh_totalincome_2000less	0.11	(0.08)	0.15	(0.07)	-0.03	(0.06)
hh_totalincome_2001to5000	0.14	(0.07)	0.16	(0.07)	0.01	(0.07)
hh_totalincome_5001to10000	0.16	(0.07)	0.15	(0.06)	-0.05	(0.05)
hh_totalincome_10001to15000	0.10	(0.05)	0.08	(0.05)	0.00	(0.04)
hh_totalincome_15001plus	0.11	(0.05)	0.08	(0.04)	-0.03	(0.04)
hh_own_tvs	-0.03	(0.02)	0.02	(0.03)	-0.03	(0.03)
hh_own_mobiles	-0.03	(0.03)	0.00	(0.02)	0.06	(0.04)
hh_own_cars	-0.02	(0.03)	0.02	(0.03)	0.02	(0.03)
hh_own_radios	0.02	(0.03)	0.03	(0.02)	0.03	(0.02)
factorwealth	0.07	(0.04)	-0.01	(0.04)	-0.07	(0.09)
hh_landown	-0.02	(0.02)	0.01	(0.03)	-0.03	(0.03)
hh_children	-0.02	(0.04)	0.13	(0.03)	0.02	(0.05)
PROVINCE1	-0.05	(0.1)	-0.25	(0.09)	0.04	(0.08)
PROVINCE2	0.20	(0.11)	-0.15	(0.11)	-0.15	(0.18)
PROVINCE3	-0.09	(0.05)	-0.14	(0.03)	0.06	(0.03)
PROVINCE4	0.03	(0.03)	0.04	(0.07)	0.14	(0.05)
PROVINCE5	0.03	(0.02)	-0.01	(0.03)	0.15	(0.04)
N	2932		2932		2932	

Table A15: Interaction Effects Regressions with Village-Level General Management Capacity Assessment [CA] Sum Scores with Agency Outcomes for Community Leaders

	CL Classroom Funds Score	
	Coef.	SE
Treatment	-0.33	(0.3)
CA sum score	0.23	(0.11)
Treatment X CA sum score	-0.26	(0.22)
Intercept	0.02	(0.15)
HAND_kapisa	1.05	(0.76)
HAND_parwan	0.38	(0.46)
int_lang_pashto	-0.05	(0.15)
hhhead_job_farmer	0.25	(0.14)
hhhead_age	-0.27	(0.12)
hhhead_schoolyears	0.08	(0.11)
school_km	0.13	(0.13)
N	123	

Table A16: Interaction Effects Regressions with Village-Level General Management Capacity Assessment [CA] Sum Scores with Agency Outcomes for Teachers

	TS Fund Application Index		TS Classroom Supplies Score		TS Motivation Index		TS Paid Regularly	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Treatment	0.26	(0.17)	-0.04	(0.19)	-0.11	(0.21)	0.39	(0.2)
CA sum score	-0.02	(0.14)	0.05	(0.11)	0.01	(0.13)	-0.07	(0.15)
Treatment X CA sum score	0.03	(0.15)	-0.02	(0.15)	0.00	(0.15)	0.07	(0.15)
Intercept	0.03	(0.15)	0.05	(0.13)	0.05	(0.16)	0.00	(0.15)
HAND_kapisa	0.22	(0.26)	0.10	(0.33)	0.11	(0.25)	0.21	(0.2)
HAND_parwan	0.38	(0.15)	0.13	(0.21)	0.26	(0.22)	0.41	(0.17)
cv_ts	-0.25	(0.1)	-0.09	(0.08)	-0.13	(0.1)	-0.27	(0.12)
N	154		154		154		154	

G Full Results for Table A11 in Main Appendix

The following Tables A17-A20 include the full regression output that corresponds with Table A11 in the main appendix on interaction effects regressions with village-level PCA scores of previous capacity building of CDCs with main child outcomes and agency outcomes for households, leaders, and teachers. For each of these tables, please note that cluster-robust standard errors are in parentheses. To construct the measure for the previous capacity building of CDCs, we use the first factor score from a principal component analysis (PCA) on ten questions on previous capacity building of CDCs.

Table A17: Interaction Effects Regressions with Village-Level PCA Scores of Previous Capacity Building of CDCs with Main Child Outcomes

	Attendance		Child Test Score	
	Coef.	SE	Coef.	SE
Treatment	-0.04	(0.04)	-0.03	(0.09)
CA PCA score	-0.01	(0.01)	0.01	(0.02)
Treatment X CA PCA score	0.01	(0.01)	-0.01	(0.02)
Intercept	0.81	(0.02)	0.75	(0.04)
HAND_kapisa	-0.02	(0.06)	0.05	(0.15)
HAND_parwan	0.11	(0.06)	0.15	(0.14)
ros_hhheadchild	0.00	(0.01)	0.05	(0.02)
ros_girl	-0.05	(0.01)	-0.08	(0.02)
ros_age	0.01	(0.01)	0.32	(0.02)
int_lang_pashto	-0.02	(0.01)	-0.01	(0.05)
hh_eth_hazara	0.02	(0.02)	0.13	(0.04)
hh_eth_pashtun	-0.04	(0.02)	-0.06	(0.05)
hh_eth_tajik	-0.02	(0.03)	0.09	(0.06)
hhhead_job_farmer	0.00	(0.01)	-0.03	(0.03)
hhhead_age	-0.01	(0.01)	0.03	(0.02)
hhhead_schoolyears	0.00	(0.02)	-0.02	(0.03)
hh_num	0.01	(0.02)	0.07	(0.03)
hh_jeribs	0.00	(0.01)	0.00	(0.02)
hh_animal_sheep	-0.01	(0.01)	-0.06	(0.03)
school_km	-0.03	(0.02)	-0.03	(0.05)
hhhead_job_laborer	-0.02	(0.01)	-0.04	(0.03)
hhhead_job_military	0.00	(0.01)	-0.03	(0.02)
hhhead_write	0.01	(0.01)	0.03	(0.04)
hhhead_read	0.02	(0.01)	0.03	(0.04)
hhhead_edu_mosque	0.02	(0.01)	0.07	(0.02)
hhhead_edu_madrassa	0.00	(0.01)	-0.03	(0.02)
hhhead_edu_community	-0.01	(0.01)	0.01	(0.01)
hhhead_edu_government	0.00	(0.01)	0.09	(0.04)
hhhead_edu_university	0.00	(0.01)	0.00	(0.02)
hhhead_female	-0.02	(0.01)	0.04	(0.03)
hh_totalincome_2000less	0.01	(0.02)	0.10	(0.06)
hh_totalincome_2001to5000	0.00	(0.02)	0.10	(0.06)
hh_totalincome_5001to10000	0.03	(0.02)	0.12	(0.06)
hh_totalincome_10001to15000	0.00	(0.02)	0.08	(0.04)
hh_totalincome_15001plus	0.00	(0.01)	0.07	(0.04)
hh_own_tvs	0.02	(0.01)	0.03	(0.03)
hh_own_mobiles	0.00	(0.01)	0.00	(0.02)
hh_own_cars	-0.01	(0.02)	0.00	(0.02)
hh_own_radios	0.00	(0.01)	-0.01	(0.02)
factorwealth	0.00	(0.02)	0.06	(0.04)
hh_landown	0.01	(0.01)	0.01	(0.02)
hh_children	-0.02	(0.01)	-0.05	(0.03)
PROVINCE1	0.00	(0.04)	-0.05	(0.07)
PROVINCE2	0.00	(0.04)	-0.09	(0.1)
PROVINCE3	0.01	(0.01)	-0.01	(0.04)
PROVINCE4	0.01	(0.02)	0.03	(0.04)
PROVINCE5	0.01	(0.01)	-0.02	(0.04)
	3894		3003	

Table A18: Interaction Effects Regressions with Village-Level PCA Scores of Previous Capacity Building of CDCs with Agency Outcomes for Households

	HH Conf. in Local Instns.		HH Perceived Teacher Quality		HH Perceived Teacher Attendance	
	Coef.	SE	Coef.	SE	Coef.	SE
Treatment	0.01	(0.08)	-0.12	(0.15)	-0.09	(0.1)
CA PCA score	0.01	(0.02)	0.01	(0.02)	0.05	(0.03)
Treatment X CA PCA score	-0.01	(0.03)	-0.05	(0.03)	-0.07	(0.04)
Intercept	0.03	(0.04)	-0.05	(0.07)	-0.06	(0.09)
HAND_kapisa	0.00	(0.13)	0.53	(0.23)	0.47	(0.31)
HAND_parwan	0.14	(0.18)	-0.04	(0.21)	0.24	(0.16)
int_lang_pashto	-0.02	(0.04)	0.05	(0.04)	0.11	(0.07)
hh_eth_hazara	0.05	(0.04)	0.02	(0.05)	0.07	(0.06)
hh_eth_pashtun	0.04	(0.06)	0.13	(0.06)	0.10	(0.06)
hh_eth_tajik	0.03	(0.08)	0.16	(0.08)	0.06	(0.09)
hhhead_job_farmer	0.03	(0.02)	0.05	(0.02)	-0.03	(0.04)
hhhead_age	0.02	(0.02)	0.01	(0.02)	0.04	(0.04)
hhhead_schoolyears	0.02	(0.03)	0.00	(0.04)	0.02	(0.05)
hh_num	0.06	(0.04)	0.04	(0.04)	-0.02	(0.04)
hh_jeribs	-0.06	(0.03)	-0.06	(0.03)	0.04	(0.03)
hh_animal_sheep	0.00	(0.03)	0.02	(0.03)	0.00	(0.04)
school_km	-0.04	(0.03)	-0.04	(0.05)	-0.02	(0.04)
hhhead_job_laborer	-0.03	(0.03)	0.02	(0.02)	-0.01	(0.03)
hhhead_job_military	-0.05	(0.03)	-0.03	(0.02)	-0.01	(0.04)
hhhead_write	0.00	(0.05)	-0.01	(0.03)	-0.09	(0.05)
hhhead_read	-0.03	(0.05)	0.04	(0.04)	0.03	(0.04)
hhhead_edu_mosque	0.06	(0.04)	0.04	(0.03)	0.07	(0.03)
hhhead_edu_madrassa	0.03	(0.03)	-0.01	(0.03)	0.00	(0.04)
hhhead_edu_community	0.03	(0.02)	0.01	(0.03)	0.00	(0.02)
hhhead_edu_government	0.00	(0.04)	-0.02	(0.04)	0.04	(0.04)
hhhead_edu_university	0.03	(0.02)	0.03	(0.02)	0.03	(0.03)
hhhead_female	0.01	(0.02)	0.00	(0.02)	-0.06	(0.03)
hh_totalincome_2000less	0.11	(0.08)	0.16	(0.07)	-0.03	(0.06)
hh_totalincome_2001to5000	0.14	(0.07)	0.16	(0.07)	0.02	(0.07)
hh_totalincome_5001to10000	0.16	(0.07)	0.16	(0.06)	-0.04	(0.05)
hh_totalincome_10001to15000	0.10	(0.05)	0.09	(0.05)	0.00	(0.04)
hh_totalincome_15001plus	0.11	(0.05)	0.09	(0.04)	-0.03	(0.04)
hh_own_tvs	-0.03	(0.02)	0.02	(0.03)	-0.03	(0.03)
hh_own_mobiles	-0.03	(0.03)	0.00	(0.02)	0.06	(0.04)
hh_own_cars	-0.02	(0.03)	0.01	(0.03)	0.02	(0.03)
hh_own_radios	0.02	(0.03)	0.04	(0.02)	0.03	(0.03)
factorwealth	0.07	(0.04)	-0.01	(0.04)	-0.07	(0.09)
hh_landown	-0.02	(0.02)	0.02	(0.03)	-0.03	(0.03)
hh_children	-0.02	(0.04)	0.13	(0.03)	0.02	(0.05)
PROVINCE1	-0.05	(0.1)	-0.26	(0.09)	0.02	(0.08)
PROVINCE2	0.20	(0.11)	-0.15	(0.12)	-0.18	(0.17)
PROVINCE3	-0.09	(0.05)	-0.14	(0.04)	0.03	(0.04)
PROVINCE4	0.03	(0.04)	0.03	(0.07)	0.11	(0.05)
PROVINCE5	0.03	(0.02)	0.00	(0.03)	0.16	(0.04)
N	2932		2932		2027	

Table A19: Interaction Effects Regressions with Village-Level PCA Scores of Previous Capacity Building of CDCs with Agency Outcomes for Community Leaders

	CL Classroom Funds Score	
	Coef.	SE
Treatment	-0.38	(0.3)
CA PCA score	0.10	(0.05)
Treatment X CA PCA score	-0.05	(0.09)
Intercept	0.02	(0.15)
HAND_kapisa	1.15	(0.73)
HAND_parwan	0.42	(0.46)
int_lang_pashto	-0.04	(0.13)
hhhead_job_farmer	0.23	(0.14)
hhhead_age	-0.28	(0.12)
hhhead_schoolyears	0.10	(0.11)
school_km	0.17	(0.13)
N	123	

Table A20: Interaction Effects Regressions with Village-Level PCA Scores of Previous Capacity Building of CDCs with Agency Outcomes for Teachers

	TS Fund Application Index		TS Classroom Supplies Score		TS Motivation Index		TS Paid Regularly	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Treatment	0.28	(0.17)	-0.01	(0.19)	-0.13	(0.21)	0.39	(0.19)
CA PCA score	-0.04	(0.06)	-0.08	(0.05)	-0.05	(0.06)	0.02	(0.06)
Treatment X CA PCA score	0.02	(0.06)	0.08	(0.07)	0.12	(0.07)	-0.04	(0.07)
Intercept	0.02	(0.15)	0.04	(0.13)	0.04	(0.16)	0.00	(0.15)
HAND_kapisa	0.19	(0.25)	0.08	(0.32)	0.16	(0.26)	0.19	(0.2)
HAND_parwan	0.37	(0.15)	0.10	(0.22)	0.24	(0.21)	0.41	(0.17)
cv_ts	-0.27	(0.1)	-0.11	(0.08)	-0.15	(0.1)	-0.27	(0.13)
N	154		154		154		154	

H Full Results for Table A12 in Main Appendix

The following Tables A21-A24 include the full regression output that corresponds with Table A12 in the main appendix on interaction effects regressions with village-level PCA scores of current CDC functionality with main child outcomes and agency outcomes for households, leaders, and teachers. For each of these tables, please note that cluster-robust standard errors are in parentheses. To construct the measure for current CDC functionality, we use the first factor score from a principal component analysis (PCA) on seven questions on previous capacity building of CDCs.

Table A21: Interaction Effects Regressions with Village-Level PCA Scores of Current CDC Functionality with Main Child Outcomes

	Attendance		Child Test Score	
	Coef.	SE	Coef.	SE
Treatment	-0.04	(0.04)	-0.02	(0.1)
CA PCA score	-0.01	(0.01)	0.03	(0.02)
Treatment X CA PCA score	0.01	(0.01)	-0.02	(0.03)
Intercept	0.81	(0.02)	0.75	(0.04)
HAND_kapisa	-0.02	(0.07)	0.04	(0.15)
HAND_parwan	0.12	(0.06)	0.14	(0.14)
ros_hhheadchild	0.00	(0.01)	0.05	(0.02)
ros_girl	-0.05	(0.01)	-0.08	(0.02)
ros_age	0.01	(0.01)	0.32	(0.02)
int_lang_pashto	-0.02	(0.01)	0.00	(0.05)
hh_eth_hazara	0.02	(0.02)	0.13	(0.04)
hh_eth_pashtun	-0.04	(0.02)	-0.06	(0.05)
hh_eth_tajik	-0.02	(0.03)	0.10	(0.06)
hhhead_job_farmer	0.00	(0.01)	-0.03	(0.03)
hhhead_age	-0.01	(0.01)	0.03	(0.02)
hhhead_schoolyears	0.00	(0.02)	-0.02	(0.03)
hh_num	0.01	(0.02)	0.07	(0.03)
hh_jeribs	0.00	(0.01)	0.00	(0.02)
hh_animal_sheep	-0.01	(0.01)	-0.06	(0.03)
school_km	-0.03	(0.02)	-0.02	(0.06)
hhhead_job_laborer	-0.02	(0.01)	-0.05	(0.03)
hhhead_job_military	0.00	(0.01)	-0.03	(0.02)
hhhead_write	0.01	(0.01)	0.03	(0.03)
hhhead_read	0.02	(0.01)	0.03	(0.04)
hhhead_edu_mosque	0.02	(0.01)	0.07	(0.02)
hhhead_edu_madrassa	0.00	(0.01)	-0.03	(0.02)
hhhead_edu_community	-0.01	(0.01)	0.01	(0.01)
hhhead_edu_government	0.00	(0.01)	0.09	(0.04)
hhhead_edu_university	0.00	(0.01)	0.00	(0.02)
hhhead_female	-0.02	(0.01)	0.04	(0.02)
hh_totalincome_2000less	0.01	(0.02)	0.10	(0.06)
hh_totalincome_2001to5000	0.00	(0.02)	0.10	(0.06)
hh_totalincome_5001to10000	0.03	(0.02)	0.12	(0.06)
hh_totalincome_10001to15000	0.00	(0.02)	0.08	(0.04)
hh_totalincome_15001plus	0.00	(0.01)	0.07	(0.04)
hh_own_tvs	0.02	(0.01)	0.03	(0.03)
hh_own_mobiles	0.00	(0.01)	-0.01	(0.02)
hh_own_cars	-0.01	(0.02)	0.00	(0.02)
hh_own_radios	0.00	(0.01)	-0.01	(0.02)
factorwealth	0.00	(0.02)	0.06	(0.04)
hh_landown	0.01	(0.01)	0.02	(0.02)
hh_children	-0.02	(0.01)	-0.05	(0.03)
PROVINCE1	0.00	(0.03)	-0.05	(0.08)
PROVINCE2	0.00	(0.04)	-0.09	(0.1)
PROVINCE3	0.01	(0.01)	-0.01	(0.04)
PROVINCE4	0.02	(0.02)	0.02	(0.04)
PROVINCE5	0.02	(0.02)	-0.02	(0.04)
N	3894		3003	

Table A22: Interaction Effects Regressions with Village-Level PCA Scores of Current CDC Functionality with Agency Outcomes for Households

	HH Conf. in Local Instns.		HH Perceived Teacher Quality		HH Perceived Teacher Attendance	
	Coef.	SE	Coef.	SE	Coef.	SE
Treatment	0.00	(0.08)	-0.15	(0.15)	-0.11	(0.11)
CA PCA score	-0.01	(0.03)	-0.02	(0.03)	-0.01	(0.03)
Treatment X CA PCA score	-0.05	(0.04)	-0.05	(0.03)	0.03	(0.05)
Intercept	0.03	(0.04)	-0.05	(0.07)	-0.07	(0.09)
HAND_kapisa	0.01	(0.12)	0.55	(0.24)	0.52	(0.33)
HAND_parwan	0.18	(0.18)	0.02	(0.2)	0.22	(0.16)
int_lang_pashto	-0.03	(0.03)	0.06	(0.05)	0.13	(0.07)
hh_eth_hazara	0.06	(0.04)	0.04	(0.05)	0.07	(0.06)
hh_eth_pashtun	0.05	(0.06)	0.13	(0.06)	0.08	(0.06)
hh_eth_tajik	0.04	(0.08)	0.17	(0.08)	0.04	(0.09)
hhhead_job_farmer	0.03	(0.02)	0.04	(0.02)	-0.03	(0.04)
hhhead_age	0.01	(0.02)	0.00	(0.02)	0.04	(0.04)
hhhead_schoolyears	0.02	(0.03)	0.00	(0.04)	0.03	(0.05)
hh_num	0.06	(0.04)	0.04	(0.04)	-0.02	(0.04)
hh_jeribs	-0.06	(0.03)	-0.06	(0.03)	0.05	(0.03)
hh_animal_sheep	-0.01	(0.03)	0.02	(0.03)	0.01	(0.04)
school_km	-0.03	(0.05)	-0.03	(0.06)	-0.02	(0.06)
hhhead_job_laborer	-0.03	(0.03)	0.01	(0.02)	-0.01	(0.03)
hhhead_job_military	-0.05	(0.03)	-0.03	(0.02)	-0.01	(0.04)
hhhead_write	0.00	(0.05)	-0.01	(0.03)	-0.09	(0.05)
hhhead_read	-0.03	(0.05)	0.04	(0.04)	0.03	(0.04)
hhhead_edu_mosque	0.06	(0.04)	0.04	(0.03)	0.07	(0.03)
hhhead_edu_madrassa	0.03	(0.03)	-0.01	(0.03)	0.00	(0.04)
hhhead_edu_community	0.03	(0.02)	0.01	(0.03)	0.00	(0.02)
hhhead_edu_government	0.00	(0.04)	-0.02	(0.04)	0.04	(0.04)
hhhead_edu_university	0.03	(0.02)	0.03	(0.02)	0.03	(0.03)
hhhead_female	0.01	(0.02)	0.00	(0.02)	-0.06	(0.03)
hh_totalincome_2000less	0.11	(0.08)	0.15	(0.07)	-0.03	(0.06)
hh_totalincome_2001to5000	0.15	(0.07)	0.16	(0.07)	0.02	(0.07)
hh_totalincome_5001to10000	0.16	(0.07)	0.16	(0.06)	-0.05	(0.06)
hh_totalincome_10001to15000	0.11	(0.05)	0.08	(0.05)	0.00	(0.04)
hh_totalincome_15001plus	0.11	(0.05)	0.09	(0.04)	-0.03	(0.04)
hh_own_tvs	-0.03	(0.02)	0.02	(0.03)	-0.03	(0.03)
hh_own_mobiles	-0.03	(0.03)	0.00	(0.02)	0.06	(0.04)
hh_own_cars	-0.02	(0.03)	0.01	(0.02)	0.02	(0.03)
hh_own_radios	0.02	(0.03)	0.04	(0.02)	0.03	(0.03)
factorwealth	0.07	(0.04)	-0.01	(0.04)	-0.06	(0.09)
hh_landown	-0.02	(0.02)	0.01	(0.03)	-0.03	(0.03)
hh_children	-0.02	(0.04)	0.12	(0.03)	0.02	(0.05)
PROVINCE1	-0.04	(0.1)	-0.24	(0.09)	0.01	(0.08)
PROVINCE2	0.21	(0.11)	-0.14	(0.12)	-0.21	(0.19)
PROVINCE3	-0.09	(0.05)	-0.13	(0.03)	0.02	(0.04)
PROVINCE4	0.02	(0.04)	0.03	(0.06)	0.11	(0.05)
PROVINCE5	0.02	(0.03)	-0.02	(0.03)	0.13	(0.04)
N	2932		2932		2027	

Table A23: Interaction Effects Regressions with Village-Level PCA Scores of Current CDC Functionality with Agency Outcomes for Community Leaders

	CL Classroom Funds Score	
	Coef.	SE
Treatment	-0.36	(0.31)
CA PCA score	0.06	(0.06)
Treatment X CA PCA score	-0.05	(0.11)
Intercept	0.04	(0.16)
HAND_kapisa	1.06	(0.79)
HAND_parwan	0.40	(0.46)
int_lang_pashto	-0.03	(0.16)
hhhead_job_farmer	0.24	(0.14)
hhhead_age	-0.27	(0.12)
hhhead_schoolyears	0.07	(0.11)
school_km	0.17	(0.13)
N	123	

Table A24: Interaction Effects Regressions with Village-Level PCA Scores of Current CDC Functionality with Agency Outcomes for Teachers

	TS Fund Application Index		TS Classroom Supplies Score		TS Motivation Index		TS Paid Regularly	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Treatment	0.29	(0.16)	0.02	(0.18)	-0.08	(0.21)	0.38	(0.19)
CA PCA score	-0.07	(0.05)	-0.12	(0.05)	-0.07	(0.06)	0.01	(0.06)
Treatment X CA PCA score	0.08	(0.06)	0.19	(0.08)	0.12	(0.08)	-0.05	(0.06)
Intercept	0.01	(0.15)	0.03	(0.12)	0.04	(0.16)	0.00	(0.15)
HAND_kapisa	0.18	(0.26)	0.02	(0.33)	0.05	(0.27)	0.23	(0.21)
HAND_parwan	0.35	(0.15)	0.06	(0.21)	0.21	(0.21)	0.43	(0.17)
cv_ts	-0.28	(0.09)	-0.13	(0.08)	-0.16	(0.1)	-0.27	(0.12)
N	154		154		154		154	

I Full Results for Table A13 in Main Appendix

The following Tables A25-A26 include the full regression output that corresponds with Table A13 in the main appendix on the effects of the sustainability model on sustainability outcomes for teachers and community leaders. Both tables include cluster-robust standard errors in parentheses.

Table A25: Effects of Sustainability Model on Sustainability Outcomes for Community Leaders

	Combined Sustain. Index		Community Provision for Future Classes		Confidence in Community Instns.		Confidence that School Will be Sustained		MOE Provisions for Funding Teachers		Shura Provisions for Future Classes	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Treatment	0.09	(0.17)	0.24	(0.18)	-0.44	(0.2)	0.08	(0.17)	-0.02	(0.19)	0.29	(0.19)
Intercept	-0.02	(0.13)	0.00	(0.12)	-0.01	(0.14)	-0.03	(0.12)	0.01	(0.14)	0.00	(0.13)
int_lang_pashto	-0.07	(0.18)	0.02	(0.19)	-0.02	(0.14)	0.00	(0.12)	-0.10	(0.02)	-0.08	(0.17)
hhhead_job_farmer	0.09	(0.08)	0.07	(0.09)	0.18	(0.09)	-0.04	(0.1)	0.03	(0.09)	0.01	(0.1)
hhhead_age	-0.27	(0.08)	-0.09	(0.09)	-0.30	(0.1)	-0.27	(0.1)	-0.05	(0.11)	-0.14	(0.09)
hhhead_schoolyears	-0.01	(0.08)	0.03	(0.1)	-0.01	(0.07)	-0.10	(0.1)	0.14	(0.12)	0.03	(0.08)
school_km	0.06	(0.1)	0.00	(0.09)	0.06	(0.1)	0.04	(0.1)	-0.03	(0.08)	-0.10	(0.09)
N	121		129		121		121		129		129	

Table A26: Effects of Sustainability Model on Sustainability Outcomes for Teachers

	Combined Sustain. Index		Confidence in Local Community Instns.		Confidence that School Will be Sustained		Proactive Management from MOE		Proactive Management from Shuras		Teacher plans to Remain CBE Teacher	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Treatment	-0.45	(0.21)	0.03	(0.04)	-0.29	(0.08)	-0.24	(0.18)	-0.01	(0.19)	-0.44	(0.19)
Intercept	0.01	(0.17)	0.04	(0.02)	0.86	(0.04)	0.00	(0.15)	0.04	(0.16)	0.00	(0.12)
ts_cv	-0.05	(0.11)	-0.02	(0.02)	0.09	(0.03)	0.02	(0.11)	-0.17	(0.08)	-0.01	(0.08)
N	154		154		154		154		154		154	

J Full Results for Table A14 in Main Appendix

The following Tables A27 corresponds to Table A14 in the main appendix on heterogeneous treatment effects by implementing NGO for the main children's outcomes and household satisfaction. Note that cluster-robust standard errors are in parentheses.

Table A27: Heterogeneous Treatment Effects by Implementing NGO

	Child Attendance			Child Test Score			Household Satisfaction		
	Coef.	SE	Int. term p-value	Coef.	SE	Int. term p-value	Coef.	SE	Int. term p-value
Treatment	-0.01	(0.03)	0.78	-0.03	(0.07)	0.81	0.04	(0.06)	0.41
CRS	0.05	(0.07)		0.08	(0.18)		-0.30	(0.2)	
Treatment X CRS	-0.01	(0.05)		0.03	(0.11)		-0.13	(0.16)	
Intercept	0.79	(0.04)		0.73	(0.09)		0.15	(0.09)	
ros_hhheadchild	0.00	(0.01)		0.06	(0.02)				
ros_girl	-0.05	(0.01)		-0.09	(0.02)				
ros_age	0.01	(0.01)		0.33	(0.02)				
int_lang_pashto	-0.02	(0.01)		-0.03	(0.04)		-0.09	(0.04)	
hh_eth_hazara	0.03	(0.02)		0.14	(0.04)		-0.02	(0.04)	
hh_eth_pashtun	-0.01	(0.02)		-0.02	(0.05)		0.06	(0.05)	
hh_eth_tajik	0.00	(0.03)		0.10	(0.06)		0.11	(0.07)	
hhhead_job_farmer	0.00	(0.01)		-0.04	(0.03)		0.02	(0.02)	
hhhead_age	-0.01	(0.01)		0.02	(0.02)		0.00	(0.03)	
hhhead_schoolyears	-0.01	(0.01)		0.00	(0.03)		0.02	(0.03)	
hh_num	0.01	(0.01)		0.05	(0.03)		-0.05	(0.04)	
hh_jeribs	0.01	(0.01)		0.01	(0.02)		-0.04	(0.02)	
hh_animal_sheep	-0.01	(0.01)		-0.04	(0.02)		0.04	(0.02)	
school_km	-0.02	(0.01)		-0.01	(0.05)		0.00	(0.05)	
hhhead_job_laborer	-0.02	(0.01)		-0.04	(0.02)		0.00	(0.03)	
hhhead_job_military	0.00	(0.01)		-0.02	(0.02)		0.00	(0.02)	
hhhead_write	0.01	(0.01)		0.03	(0.03)		0.00	(0.04)	
hhhead_read	0.02	(0.01)		0.05	(0.03)		-0.02	(0.04)	
hhhead_edu_mosque	0.02	(0.01)		0.04	(0.02)		-0.03	(0.03)	
hhhead_edu_madrassa	-0.01	(0.01)		-0.01	(0.01)		0.00	(0.02)	
hhhead_edu_community	0.00	(0.01)		0.01	(0.01)		0.03	(0.02)	
hhhead_edu_government	0.00	(0.01)		0.07	(0.04)		0.02	(0.03)	
hhhead_edu_university	0.00	(0.01)		0.00	(0.02)		-0.01	(0.03)	
hhhead_female	-0.01	(0.01)		0.04	(0.02)		0.02	(0.02)	
hh_totalincome_2001less	0.02	(0.02)		0.07	(0.05)		0.02	(0.07)	
hh_totalincome_2001to5000	0.01	(0.02)		0.07	(0.04)		0.03	(0.06)	
hhhead_totalincome_5001to10000	0.02	(0.02)		0.07	(0.04)		0.08	(0.06)	
hhhead_totalincome_10001to15000	0.01	(0.02)		0.05	(0.03)		0.01	(0.04)	
hh_totalincome_15001plus	0.01	(0.01)		0.04	(0.03)		0.04	(0.03)	
hh_own_tvs	0.03	(0.01)		0.03	(0.02)		0.02	(0.02)	
hh_own_mobiles	0.01	(0.01)		0.00	(0.02)		0.04	(0.02)	
hh_own_cars	0.00	(0.01)		0.01	(0.02)		-0.01	(0.02)	
hh_own_radios	0.01	(0.01)		-0.01	(0.02)		0.01	(0.02)	
factorwealth	-0.01	(0.02)		0.06	(0.04)		0.01	(0.03)	
hh_landown	0.00	(0.01)		0.01	(0.02)		0.00	(0.02)	
hh_children	-0.02	(0.01)		-0.05	(0.02)		0.10	(0.03)	
PROVINCE1	0.04	(0.01)		0.02	(0.03)		-0.22	(0.03)	
PROVINCE3	0.00	(0.01)		0.00	(0.04)		-0.09	(0.04)	
PROVINCE4	0.02	(0.01)		0.03	(0.03)		0.08	(0.06)	
PROVINCE5	0.02	(0.01)		-0.02	(0.04)		0.02	(0.04)	
N	4980			3822			3750		

K Full Results for Table A15 in Main Appendix

The following Tables A28-A29 corresponds to Table A15 in the main appendix on community leaders' longer-term perceptions of CBE sustainability, with both the CARE and CRS samples. Note that cluster-robust standard errors are in parentheses.

Table A28: Community Leaders' Longer-Term Perceptions of CBE Sustainability (CARE and CRS Samples) - Equivalency Outcome

	Community Desire to Continue CBE	
	Coef.	SE
Treatment	-0.03	(0.02)
Intercept	1.00	(0)
int_lang_pashto	0.00	(0)
hhhead_job_farmer	0.00	(0.01)
hhhead_age	-0.03	(0.02)
hhhead_schoolyears	0.01	(0)
school_km	0.01	(0.01)
N	129	

Table A29: Community Leaders' Longer-Term Perceptions of CBE Sustainability (CARE and CRS Samples) - Non-Equivalency Outcomes

	Combined Sustainability Index		Initiative to have CBE Sustained		Provisions for Sustain. Without NGO		Confidence that School will Sustain		Continued CBE Opts. in Spring 2018	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Treatment	0.16	(0.2)	0.59	(0.19)	0.28	(0.22)	-0.13	(0.09)	-0.33	(0.09)
Intercept	-0.01	(0.13)	0.01	(0.14)	-0.01	(0.13)	0.51	(0.06)	0.60	(0.07)
int_lang_pashto	0.16	(0.14)	0.12	(0.12)	0.14	(0.15)	0.05	(0.05)	0.01	(0.06)
hhhead_job_farmer	0.02	(0.11)	0.10	(0.1)	0.05	(0.11)	-0.07	(0.05)	0.00	(0.04)
hhhead_age	0.02	(0.09)	-0.24	(0.11)	0.18	(0.09)	0.07	(0.04)	0.01	(0.04)
hhhead_schoolyears	-0.13	(0.11)	0.02	(0.08)	-0.24	(0.09)	0.01	(0.05)	0.00	(0.04)
school_km	0.17	(0.1)	-0.10	(0.09)	0.09	(0.08)	0.01	(0.05)	0.12	(0.05)
N	129		129		123		129		129	

L Full Results for Table A16 in Main Appendix

The following Tables A30-A31 corresponds to Table A16 in the main appendix on interaction effect regressions with community leaders' age and educational status for the endline community leader sustainability perception outcomes. Note that cluster-robust standard errors are in parentheses. In the first table, "older leader" is defined as having an age above the median for community leaders (52 in our sample). In the second table, "educated" means having completed any formal education (39% of leaders in our sample).

Table A30: Interaction Effect Regressions With Community Leaders' Age, Endline Community Leader Sustainability Perceptions

	Combined Sustain. Index		Community Provision for Future Classes		Confidence in Community Instns.		Confidence that School Will be Sustained		MOE Provisions for Funding Teachers		Shura Provisions for Future Classes	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Treatment	0.32	(0.24)	0.53	(0.26)	0.14	(0.25)	0.19	(0.26)	-0.18	(0.27)	0.13	(0.26)
Older leader	0.15	(0.32)	0.86	(0.41)	-0.01	(0.32)	-0.10	(0.34)	-0.41	(0.33)	0.50	(0.34)
Treatment x Older Leader	-0.45	(0.31)	-0.61	(0.38)	-1.09	(0.34)	-0.20	(0.38)	0.33	(0.3)	0.28	(0.34)
Intercept	-0.09	(0.19)	-0.41	(0.21)	0.01	(0.21)	0.02	(0.21)	0.21	(0.26)	-0.24	(0.2)
int_lang_pashto	-0.07	(0.18)	-0.01	(0.17)	-0.02	(0.11)	0.00	(0.13)	-0.09	(0.03)	-0.09	(0.16)
hhhead_job_farmer	0.11	(0.08)	0.08	(0.09)	0.22	(0.08)	-0.03	(0.1)	0.02	(0.09)	-0.01	(0.09)
hhhead_age	-0.24	(0.13)	-0.32	(0.17)	-0.08	(0.17)	-0.19	(0.16)	0.05	(0.18)	-0.39	(0.15)
hhhead_schoolyears	-0.01	(0.08)	0.03	(0.1)	-0.01	(0.06)	-0.10	(0.09)	0.14	(0.12)	0.03	(0.08)
school_km	0.06	(0.1)	-0.02	(0.1)	0.08	(0.1)	0.05	(0.1)	-0.02	(0.07)	-0.12	(0.08)
N	121		129		121		121		129		129	

Table A31: Interaction Effect Regressions With Community Leaders' Educational Status, Endline Community Leader Sustainability Perceptions

	Combined Sustain. Index		Community Provision for Future Classes		Confidence in Community Instns.		Confidence that School Will be Sustained		MOE Provisions for Funding Teachers		Shura Provisions for Future Classes	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Treatment	0.09	(0.22)	0.34	(0.24)	-0.49	(0.25)	-0.02	(0.24)	-0.18	(0.27)	0.27	(0.24)
Intercept	-0.07	(0.17)	-0.03	(0.16)	0.10	(0.19)	-0.07	(0.16)	0.21	(0.26)	-0.10	(0.15)
Educated leader	0.12	(0.26)	0.08	(0.27)	-0.28	(0.28)	0.11	(0.25)	-0.41	(0.33)	0.28	(0.28)
Treatment x Educated Leader	-0.02	(0.31)	-0.27	(0.35)	0.13	(0.36)	0.23	(0.36)	0.33	(0.3)	0.03	(0.38)
int_lang_pashto	-0.07	(0.19)	0.03	(0.19)	-0.02	(0.13)	0.00	(0.12)	-0.09	(0.03)	-0.08	(0.18)
hhhead_job_farmer	0.11	(0.09)	0.06	(0.09)	0.15	(0.09)	0.00	(0.11)	0.02	(0.09)	0.05	(0.11)
hhhead_age	-0.26	(0.08)	-0.10	(0.09)	-0.31	(0.11)	-0.25	(0.1)	0.05	(0.18)	-0.11	(0.09)
hhhead_schoolyears	0.00	(0.09)	0.02	(0.1)	-0.02	(0.07)	-0.08	(0.1)	0.14	(0.12)	0.05	(0.08)
school_km	0.07	(0.1)	-0.01	(0.1)	0.03	(0.11)	0.06	(0.1)	-0.02	(0.07)	-0.07	(0.09)
N	121		129		121		121		129		129	

M Full Results for Table A17 in Main Appendix

The following Tables A32 corresponds to Table A17 in the main appendix on interaction effect regressions with community mean children's test score for the endline community leader sustainability perceptions outcomes. Note that cluster-robust standard errors are in parentheses.

Table A32: Interaction Effect Regressions With Community Mean Children's Test Score, Endline Community Leader Sustainability Perceptions

	Combined Sustain. Index		Community Provision for Future Classes		Confidence in Community Instns.		Confidence that School Will be Sustained		MOE Provisions for Funding Teachers		Shura Provisions for Future Classes	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Treatment	-0.14	(0.3)	0.11	(0.28)	-0.76	(0.29)	-0.05	(0.29)	0.06	(0.25)	0.06	(0.25)
Intercept	0.22	(0.23)	0.04	(0.2)	0.37	(0.2)	0.17	(0.2)	0.00	(0.21)	-0.04	(0.18)
Mean Test Score	-0.64	(0.42)	-0.10	(0.36)	-1.03	(0.37)	-0.53	(0.4)	0.02	(0.45)	0.12	(0.45)
Treatment x Mean Test Score	0.65	(0.54)	0.36	(0.51)	0.91	(0.56)	0.38	(0.54)	-0.21	(0.5)	0.65	(0.55)
int_lang_pashto	-0.08	(0.21)	0.03	(0.19)	-0.04	(0.14)	-0.01	(0.13)	-0.11	(0.03)	-0.05	(0.16)
hhhead_job_farmer	0.09	(0.09)	0.06	(0.09)	0.15	(0.09)	-0.05	(0.1)	0.03	(0.09)	0.00	(0.09)
hhhead_age	-0.34	(0.08)	-0.10	(0.09)	-0.36	(0.1)	-0.30	(0.1)	-0.05	(0.12)	-0.14	(0.09)
hhhead_schoolyears	-0.03	(0.09)	0.02	(0.11)	-0.03	(0.07)	-0.11	(0.1)	0.14	(0.12)	0.01	(0.08)
school_km	0.05	(0.1)	0.00	(0.09)	0.05	(0.09)	0.04	(0.09)	-0.04	(0.08)	-0.07	(0.08)
N	121		129		121		121		129		129	