
Heather Huntingon, University of Michigan and Cloudburst Group
Erik Wibbels, Duke University
Brett Gall, Duke University

JULY 29, 2017

DISCLAIMER
The authors' views expressed in this publication do not necessarily reflect the views of the United States Agency for International Development or the United States Government.
<table>
<thead>
<tr>
<th>ACRONYMS AND ABBREVIATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIC</td>
</tr>
<tr>
<td>ATE</td>
</tr>
<tr>
<td>CO</td>
</tr>
<tr>
<td>COP</td>
</tr>
<tr>
<td>COR</td>
</tr>
<tr>
<td>CSO</td>
</tr>
<tr>
<td>DA</td>
</tr>
<tr>
<td>GAS</td>
</tr>
<tr>
<td>G-SAM</td>
</tr>
<tr>
<td>IE</td>
</tr>
<tr>
<td>IRB</td>
</tr>
<tr>
<td>ITT</td>
</tr>
<tr>
<td>M&amp;E</td>
</tr>
<tr>
<td>MDES</td>
</tr>
<tr>
<td>MMMDA</td>
</tr>
<tr>
<td>PAC</td>
</tr>
<tr>
<td>PFM</td>
</tr>
<tr>
<td>SSA</td>
</tr>
<tr>
<td>TOT</td>
</tr>
<tr>
<td>USAID</td>
</tr>
<tr>
<td>USG</td>
</tr>
</tbody>
</table>
The USAID Ghana Strengthening Accountability Mechanisms (G-SAM) program is a 4-year intervention designed to strengthen unaccountable or weakly accountable Metropolitan, Municipal, or District Assemblies (MMDAs). The GSAM is being implemented nationwide in Ghana, across 100 districts, from early 2015 through the end of 2018. The G-SAM program represents a ten million-dollar project that is supported by USAID/Ghana’s Democracy and Governance Office. The objective of the program is to improve local governance in Ghana through strengthening both top-down and bottom-up accountability. The work addresses USAID/Ghana’s objectives of improving governance and district-built capital projects and related service delivery, improving the capacity of Ghanaian institutions, and strengthening local democratic procedures and citizen engagement.

The impact evaluation (IE) focuses on two GSAM interventions: (1) a series of top-down performance audits by the central government’s Ghana Audit Service (GAS) of District Assembly (DA)-contracted and constructed development projects and (2) a bottom-up civil society organization (CSO)-led campaign also focused on the quality of district development projects. Citizen outreach to diffuse the findings pertaining to each respective district’s top-down or bottom-up programming has followed the top-down and bottom-up audits.

Both the top-down and the bottom-up treatment arms focus on auditing districts’ “development projects.” These projects include the schools, public toilets, clinics, feeder roads, irrigation projects, and other infrastructure and centers of public service provision that typically consume the lion’s share of district government spending. Throughout this report we refer to these district-built capital projects as “development projects.”

When these district development projects are planned and executed with citizen needs in mind, they can impart substantial developmental benefits for citizens. The government of Ghana, the USAID mission, and citizens all expressed concern that many district governments were building projects inefficiently and without citizen input. For example, there is evidence that more than 40 percent of projects are never completed (Williams 2016). In light of these concerns, the program was designed to increase accountability for the planning and execution of these crucial district government projects.

The IE is designed as a randomized control trial to assess the two different activities of the governance program. The evaluation rigorously assesses the impact of performance audits and CSO-led social audits on improving accountability and service delivery outcomes by randomly assigning one of the two activity treatments or a pure control group status to 150 districts in Ghana. Specifically, 50 districts received performance audits from the Ghana Audit Service and subsequent dissemination of the performance audit results through meetings with district officials and a public scorecard campaign. 50 similar, additional districts have received social audits from civil society organizations and subsequent dissemination of the social audit results through a different scorecard-based information dissemination campaign.

Additionally, another 50 similar districts serve as a pure control group and neither received audits through the programming nor were subject to any information dissemination campaign. As designed, this is one of the largest IEs ever conducted on a USAID governance project at the district level and one of the first to explicitly compare top-down and bottom-up interventions for promoting accountability in district governance.

Prior to the rollout of project activities, baseline data collection was carried out in each of the 150 districts. This data collection effort sought to gather information on baseline conditions on the key outcomes of interest. The endline data collection - and therefore the primary data source for the endline evaluation - is also focused on similar outcomes. These primary outcomes of interest in the endline
evaluation include:

1. The quality of district development projects and corresponding service access;
2. The quality of district governance, with special reference to district capital projects;
3. Accountability of district officials and citizen engagement with district-level democratic procedures; and
4. Awareness of GSAM and its impact on citizen information and engagement with district government.

This document describes the endline analysis plan for the GSAM IE. The plan serves as an important guide for assessing the rigor and validity of the final analysis; the authors will complete and register the plan prior to the analysis of endline data. As of the registration of this plan, field work was completed, the data were in the possession of the IE team, data were aggregated into a single file for each survey target population, and basic diagnostics were conducted for checks of data quality, but data had yet to undergo any analysis.

The pre-analysis plan is structured as follows: Sections 1 and 2 provide the research motivation and background. Section 3 presents details of the experimental design. Sections 4 and 5 describe the data and indicators used for hypotheses testing. Sections 6 and 7 outline the empirical strategy, and Section 8 covers the ethical considerations.
1. BACKGROUND

In 1988, Ghana launched its comprehensive local government and public administration reform, via six pillars of decentralization: political, administrative, fiscal, planning, market, and spatial. The reform sought to transfer power and competence to a local government system in which Metropolitan, Municipal and District Assemblies (MMDAs) consist of assembly members who are either elected to office or appointed to office. There are two types of key actors within the MMDAs: elected District Assemblies (DAs) which are intended to represent local citizens, and a District Chief Executive (DCE) which is intended to represent central authorities. While the former are elected, the President of the Republic of Ghana appoints the latter. Although “the MMDAs are nominally non-partisan, as candidates are not sponsored by political parties”, presidents have historically appointed members from their own parties (Ayee and Dickovick, 2010). A close analogy in scale to MMDAs would be county-level governments in the United States. As of April 2012, two hundred and twelve MMDAs existed in Ghana.

In Ghana, MMDAs are the only substantively meaningful level of government below the national administration. The MMDAs are primarily responsible for improving service delivery outcomes through “local-level policy and planning, [which is] based on, and informed by, national policies and programs” (Ayee and Dickovick, 2010). MMDAs rely on several sources of revenue to fund local government policies and activities. The District Assemblies' Common Fund (DACF) serves as the most important source of revenue and provides intergovernmental transfers to MMDAs. Funds from this source are distributed according to a revenue sharing formula. Additionally, Internally Generated Funds (IGFs) are derived from taxes and levies imposed by LGs. Further sources of revenue include the District Development Fund (DDF) and rents from local extractive industries (Ayee and Dickovick, 2010).

Advocates of the decentralization of Ghanaian governance have argued that decentralization can promote participatory governance, improve service delivery, and facilitate socio-economic development (Ohene-Konadu 2001). Consistent with decentralization initiatives in other countries, reforms have been guided by the belief that local governments are better able to respond to citizen needs and more accountable to citizens due to their proximity to the citizenry. The expectation has been that well-managed local governments are an important means to achieve democracy, “good governance”, and the efficient mobilization and allocation of scarce resources to meet citizens’ needs.

Ghana’s ambitious decentralization reform process and various donor initiatives notwithstanding, the MMDA system is failing to achieve its essential mandate of efficient, effective, and equitable local service delivery (Ayee and Dickovick 2010, CDD-Ghana 2008, Fox et al. 2010). Decentralization reforms have failed to address many of the basic problems associated with poor governance and service provision at the local level. In many cases, the devolution of power and responsibilities to local governments has failed to yield more efficient, effective and equitable local service delivery outcomes. Corruption, rent-seeking, and low effort provision continue to characterize some local governments. These characteristics combine to facilitate the capture of decentralized politics by local elites and interest groups, distort democratic processes and obfuscate lines of accountability. In the absence of accountability, local elected officials and technocrats face weak incentives to address the needs of broad groups of local citizens. This state of affairs has significant negative implications for the quality of local public services.

Some of the underlying challenges to accountability result from structural factors that are all but impossible to address with USAID programming. For instance, the fact that DCEs are very powerful and appointed by the president profoundly limits their incentives to respond to district citizen concerns and receive input from citizens. Since DCE promotion involves returning to the national government, they are incentivized to respond to the national party leadership. Due to the institutional incentives of DCEs, USAID programming is unlikely to produce substantial improvement in district governance.
accountability, and development projects through interventions directly aiming to influence the behavior of the DCEs.

Other gaps in local accountability, however, are amenable to improvement through USAID programming. GSAM focuses its efforts on two areas with considerable promise for improvement: (1) limited central government oversight of the MMDAs and (2) limited availability of information to citizens on MMDA budgets, development projects and service delivery.

The opportunity to improve central government oversight of the MMDAs arises from clear shortcomings with the existing means of central government monitoring of district governments. More specifically, although the law requires that GAS conduct a financial audit of all MMDA financial accounts for each fiscal year, the system of auditing is a limited tool for promoting accountability. First, the audits are only intended to identify accounting irregularities and do not provide information on the quality of roads, irrigation projects, or any other public expenditure. Thus, the existing financial audits provide no information on the quality or efficiency of MMDA expenditures. Second, the sanctioning regime for poor audit performance is very weak. Audit results first go to the DCE and the District Assembly’s Audit Report Implementation Committee (ARIC), which are free to ignore any violations. While all MMDA audits are later aggregated in a report that is delivered to the Auditor General and from there to the Parliament’s Public Accounts Committee (PAC), neither the GAS, the Auditor General, nor the PAC currently has the capacity to formally sanction MMDAs whose audits provide evidence of corruption in public financial management (PFM). The PAC can rely on public shaming and the press has taken up the cause in a few cases, but only a small number of MMDA budget violations are ultimately sanctioned or face public scrutiny.

These deficits in central government oversight produce several development failures. First, many MMDAs have weak internal PFM. Second, there is generally a weak link between revenues and the quality of public services. Third, there is substantial room for corruption in the management of MMDA budgets through mechanisms other than explicit financial fraud detectable and sanctionable under the financial audit system, and citizens possess tenuous means of discovering failures in MMDA budget management. Since MMDA budgets are overwhelming oriented toward building development projects, they represent an important tool for development, and their mismanagement represents a failure of accountability. The program design builds upon prominent recent programming and research showing powerful positive effects of audits on reducing local corruption (Olken 2007) and increasing local electoral accountability (Ferraz and Finan 2008). When properly conducted, performance audits provide information on the management of public finances and the value provided by public services.

Second, local citizens possess limited information on MMDA budgets and their link to development projects and the quality of local public services. For many of the reasons outlined above, it is difficult for citizens to gather information on the size or management of MMDA budgets. In many districts, the weakness of Assembly-level oversight of the budget further exacerbates this problem. Internal auditing procedures are compromised by weak capacity and poor system design. The Assemblies have ARICs, but heavy representation of DCE interests (including the DCE him/herself) and limited means of public outreach compromise the work of the ARIC. The confluence of all of these factors has led to an MMDA citizenry that is poorly informed with regard to district budget size and project management. As such, citizens have few means to assess the allocation or quality of district development projects in light of the budget. This implies limited mechanisms by which citizens can develop objective assessments of the value-for-money they receive from MMDA services. Moreover, since citizens typically do not observe the quality and efficiency of schools, health clinics or irrigation projects in other localities or districts, it is difficult for them to know which communities are receiving quality development projects and what constitutes relatively good or bad public services. In addition to citizens having limited opportunities to learn about relative district project and service quality through personal experience with the project
outcomes and services of other districts, citizens also face serious coordination problems in gathering and assessing such information when it can be found at all. All of these challenges were reflected in baseline household surveys, which showed limited engagement with, and considerable dissatisfaction with district governance.

Several development failures emerge from these weaknesses: MMDAs are unaccountable or weakly accountable for the quality of development projects as citizens are unsure which level of government is responsible for which public services; and citizens participate in MMDA politics at fairly low levels. Knowledge and participation are crucial mechanisms for generating good governance and low levels of either threats accountability of district governments for both good and bad outcomes. The project design builds on prominent recent randomized trials that examine the effect of bottom-up social accountability mechanisms for the quality of local public services. Andrabi, Das, and Khwaja (2009), for instance, report on a treatment that provides details on school performance, which has measurable effects on both learning and school fees. Duflo, Hanna, and Ryan (2012) and Muralidharan and Sundararaman (2011) target service providers directly and show that even a moderate improvement in incentives can induce improvements in teacher attendance. Björkman and Svensson (2009) examine the effects of community-based monitoring on local health clinics in Uganda and report impressive improvements in service provision, utilization, and health outcomes. All of these interventions are aimed at improving social accountability by providing citizens with the information necessary to make better-informed demands on the government.

In summary, the overarching development problem that this program aims to address is that MMDA officials are unaccountable. They are unaccountable to the central government, which funds the largest share of MMDA budgets. They are also unaccountable to local citizens, who are the most important beneficiaries of district-built development projects. Building on cutting edge development projects elsewhere, the impact evaluation of this program involves the randomized assignment of two main activities – performance audits and a scorecard campaign – along with the randomization of control group status to a third group of districts. Given this design, the impact evaluation allows the IE team to assess the effectiveness of each activity.

Importantly, the interventions provided information to citizens on the performance of their DAs specifically with regard to the management of district development projects. This information included (1) whether projects were planned in a participatory manner, (2) the degree of competitiveness and reliance on merit in project procurement, (3) whether the projects were completed, (4) whether project construction was well-executed so as to produce high-quality infrastructure, and (5) the extent to which the projects benefited local citizens. Where DAs performed well, citizens received positive information regarding the performance of their district. Where DAs performed poorly, citizens received negative information regarding the performance of their district. This raises the possibility, discussed in greater detail below, that the program produced heterogeneous effects such that the program improved citizen attitudes toward district governance in some districts while decreasing support for district governance in others.
2.0 INTERVENTIONS AND HYPOTHESES
GSAM implementation began in 2015 and is expected to take place over a 4-year period, across a total of 150 districts in all ten regions of Ghana. The endline report will be developed prior to the end of the program in order to accommodate specific features of program timing. The two activities under evaluation include: (1) GAS Performance Audits and corresponding citizen information campaigns; and (2) a CSO-led social audit and information campaign. The impact of each of these activities will be identified by a randomized control trial; Fifty districts received GAS performance audits, and 50 districts received the CSO scorecard campaign. An additional 50 districts serve as a control group during the period of the impact evaluation.

2.1 INTERVENTION 1: GAS PERFORMANCE AUDIT & INFORMATION CAMPAIGN
As part of G-SAM, the GAS implemented performance audits of the DAs to strengthen hierarchical accountability across levels of government. Although the GAS is legally mandated to conduct financial audits of all MMDAs, financial audits provide limited information on the quality of local governance because they do not address the quality of local public services or capital spending. Performance audits, on the other hand, are explicitly aimed at evaluating the value-for-money in public budgeting. They go well beyond accounting to cover each step in the process of planning and building development projects, including: (1) the identification and budgeting for projects in district budgets, (2) the tendering and award of contracts, (3) the monitoring, evaluation, and payment of contractors as projects are being built, and (4)
post-completion assessments of project quality and the impacts of projects on citizen welfare. Although the GAS had historically conducted performance audits of central ministries prior to the intervention, it had neither staff nor capacity to do so at the district level. These capacities were developed in advance of the program’s performance audits courtesy of G-SAM funding.

For G-SAM, GAS conducted performance audits of the top two most expensive school-building projects as reported in Annual Progress Reports of each of the 50 districts assigned to the GAS audit experimental condition. The choice was made to focus on schools as they are easily the most common district-built project, and focusing on schools allowed for a simplification of the training of auditors. These projects include classroom block construction and other infrastructural projects. The purpose of the performance audits was to ascertain whether the selected DAs are capable of selecting, planning, procuring contractors, and supervising and monitoring capital project construction, budgeting, contracting, and quality in accordance with sound administrative practices and in a way that is responsive to local needs.

The performance audits were implemented as follows. The GAS submitted notifications to 50 randomly selected DAs — including the DCE, members of the Assembly, the budget and development committees, and the District Coordinating Director — informing them that they would each receive a performance audit. The letter explained that performance audits are much more involved than standard GAS financial audits. A typical audit took two weeks and involved a combination of interviews, document review, and field work. The performance period reviewed covered a three-year period, 2012-2014. Following completion of the performance audits, GAS conducted exit meetings with district officials to provide preliminary feedback and findings. Thereafter, the GAS produced audit reports. All of the performance audits were completed by late spring 2015.

Thereafter, GAS and implementing partner CARE International disseminated results to DA officials and citizens. This began with the construction of district scorecards on the basis of the audit results (see Appendix). Through December, 2015 and January, 2016, the evaluation team worked closely with GAS and CARE International (the contractor for the dissemination of the information to district officials and citizens) on how to score the audit reports with an eye to the development of clear and accessible scorecards. This resulted in a sensible and simple way of scoring the reports. CARE and GAS scored the reports separately and then compared notes in order to ensure that the scores reflected GAS’ intent. The resulting scorecards are quite straightforward and include both objective scores and information on how the district did compared to others. The scorecards include simple graphics for those citizens that are illiterate and/or innumerate.

To present the findings and scorecards, GAS and CARE conducted widely publicized, public information-sharing sessions with DA officials and invited citizens and CSOs in all 50 district capitals during April and May of 2016. This timing proved a serious difficulty for the midline survey team (see below), but the GAS/CARE scorecard events appear to have gone well. It seems DA officials were appropriately concerned about the findings, and some citizen groups reported that this was the first time they had heard that they could complain about incomplete and poorly constructed projects. The schedule

---

1 These reports are required and their filing is a precondition for receiving the following year’s fiscal transfers from the central government. They include information on each project, including project type, location, funding source and level of completion.

2 The inclusion of water or electrification projects, for instance, have quite different engineering challenges; thus, assessments of project quality are quite different than for buildings like schools.

3 See Annex 1 for an example of one GAS’ comprehensive performance audit reports.

4 See Annex 2 for an example of the scorecards.
of these events is available upon request.

Lastly, a large-scale public information and mobilization campaign followed the series of DA meetings in the GAS treated districts, which included: a) scorecard dissemination events in 25 communities per GAS district; b) dissemination events in “alternative spaces” to reach individuals who are often excluded from formal events, including youth, women and those with disabilities; and c) radio jingles to distribute audit results and encourage project monitoring. Scorecards were distributed by CARE International and their partners to DA citizens at public events in each of the 50 GAS audit districts. These dissemination events occurred in 25 communities in each GAS district for a total of 1250 meetings.\(^5\) In addition to the distribution of scorecards, the meetings also involved the provision of information regarding the DA and which DA officials are responsible for project quality, such as the identities of the district representative, members of the relevant DA committees, and the district administrators. CSOs also engaged with DA citizens to demand accountability for the results. These 1,250 local meetings were facilitated by the coordinating CSOs and district-level community-based organizations (CBOs), provided a forum for discussing the scorecards and mobilizing citizens around the results. This hands-on approach was complemented by efforts to diffuse the results via print and radio.

The dissemination events that occurred in “alternative spaces” were aimed at typically marginalized populations and took place in schools, mosques, churches, markets, and taxi stands. These events covered similar material to that discussed above; there were 123 such events reaching 12,099 citizens across the 50 GAS districts. Finally, the radio jingles aired GAS audit results, encouraged citizen participation in oversight of development projects, and provided phone-in segments for public discussion.\(^6\)

The GAS program activities could have feasibly influenced DA planning, budgeting, and project oversight through two mechanisms. First, the performance audits themselves may have promoted changes in district practices. Indeed, GAS, CARE, and our own team noted that DA officials were nervous about the arrival of the GAS in their districts and about the content of the initial reports as conveyed in the exit interviews. Second, it could be that citizen attention to audit results is a prerequisite for behavioral change in the DA. In those DAs that receive positive reports, it is expected that the campaign will improve citizen attitudes about the quality of local governance. In those DAs that receive negative reports, it is expected that citizens will demand improved projects and services and demand accountability. Given the relatively small number of districts involved in this arm of programming, our capacity to detect these heterogeneous effects is limited; we discuss this challenge in greater detail below. In any case, the programming provided the 50 DAs with stronger incentives to perform well in the face of an increase in the probability of facing sanctions due to the outcomes of the performance audit by providing central government officials and local citizens with timely information on the audit outcomes. With improved information, both central government officials and citizens should be empowered to hold district governments more accountable for their management of the public purse.

2.2. INTERVENTION 2: CSO-LED SOCIAL AUDIT & INFORMATION CAMPAIGN

The CSO-led social audit campaign was implemented in an additional 50 districts. The CSO social audit was designed to strengthen social accountability by providing DA citizens with detailed information - collected via the social audits - on the quality of district capital projects. This portion of the programming was contracted to CARE International. CARE led a consortium of implementing partners built on its preexisting network of 245 CSOs, over 10,388 farmers’ groups, youth groups, and 2,400 village saving and loan associations (VSLAs). These networks and relationships provided the on-the-ground

\(^{5}\) 1,185 of these meetings had occurred by the end of the third quarter of 2016; 53 additional meetings occurred in the fourth quarter of 2016; the final 12 meetings occurred in the first quarter of 2017.

\(^{6}\) As of March 2017, this radio-based dissemination had occurred in 45 of 50 GAS districts. The five where they had not occurred are Biakoye, South Dayi, Central Tongu, Asuogyaman, and North Tongu.
infrastructure for sustained social audits involving community-gathered information on the planning, passage, implementation, and use of district development budgets.

The social audit programming has four key components: First, the formation of District Steering Committees, typically with seven members, who are responsible for overseeing GSAM activities in the district, promoting citizen participation in project budgeting and monitoring, and engaging with DA officials on capital budgeting and project oversight. Second, as in the GAS districts, two capital projects per district were selected from the 2015/16 Annual Development Plans for extensive auditing. In these cases, CARE International and CSO partners trained community members in project oversight, emphasizing work quality, absenteeism by contract, the pace of work, the quality of work materials, and names, phone numbers and procedures for registering complaints with DA officials about projects. These trainings identified Community Development Monitors who helped lead the subsequent citizen auditing of projects on a biweekly basis by both CSOs and citizens. Third, the results of these social audits were aggregated into scorecards modeled on those developed for the GAS audits. The construction of the scorecards resulted from a combination of community meetings (which produced an initial “community scorecard”) and CSO monitoring data. Fourth, the resulting citizen scorecard was disseminated across the district in similar ways to those described above in the GAS districts, albeit less extensively. As in the GAS districts, dissemination has occurred via district-level town hall meetings (all 50 SA districts), community meetings (676 of them across the 50 SA districts, as compared to 1250 in GAS districts), and community radio/jingles. Unlike in the GAS districts, social audit results have been disseminated via billboards in 47 of the 50 SA districts. Dissemination events have not taken place in “alternative spaces” in the SA districts.

The local training for the social audits was scheduled to begin in the second quarter of 2015. Due to the difficulties in building and coordinating CSO capacity in so many districts, the programming was delayed until the last quarter of 2015. This delay had two significant implications for the impact evaluation. First, it made a midline evaluation of the CSO-led social audit campaign moot since too few activities had taken place. Second, the delay placed considerable time pressure on the CSOs to complete their social audit activities so that the requisite information would be in the hands of citizens in a timely manner that would allow for feedback into annual DA budgeting processes.

Despite these difficulties, brief fieldwork by the evaluation team in three districts in March 2016 and subsequent quarterly reports by CARE International indicate that the SA programming met most of its targets before the endline data collection went into the field. Social audits took place of all 100 projects in the 50 SA districts, and CARE and partners generated scorecards for all 50 districts. District-level town hall meetings have taken place in all 50 SA districts, 46 of which occurred before endline data collection began. Community-level dissemination was less widespread than in GAS districts, but reached 27,278 citizens in 613 communities across the 50 SA districts; most of these community meetings occurred in the same quarter as endline data collection, so it is unclear if the evaluation team will capture their effect. Likewise, the construction of billboards occurred during the quarter of data collection.

2.3 Intervention Timing, Implementation Challenges and Cautions: Implications for the Evaluation

Complications in program implementation had three important impacts on the IE team’s capacity to uncover the effects of the interventions. First, the delay in the social audits created a mismatch in the timing between the two arms of programming. While the GAS audits took place in the latter half of 2015 and the corresponding citizen outreach took place in the second quarter of 2016, the social audits were...

---

7 The project completion rates in these SA communities was 56%, which compares favorably to the rate of 45% in 2015 for all districts.
8 9 SA districts had not received this radio programming by March 2017.
conducted throughout 2016 and the resulting scorecards and broader outreach efforts took place in the latter half of 2016. This posed a challenge in that the IE team was unable to delay endline data collection until the realization of the complete effect of the implementation of the social audit programming. In particular, the social audits involve considerable citizen training and engagement, and this intervention is expected to continue for a further two years. Nevertheless, further delays in the endline data collection would have meant that a considerable amount of time would had passed since the GAS performance audits. The team’s capacity to detect the most pronounced short-term and medium-term effects of the GAS programming would be limited by a two to three year wait for the social audit programming to take effect. As a result, it could be that the evaluation will underestimate the social audit intervention’s longer term effects.

Second, while both interventions were initially conceptualized as district-wide interventions insofar as they involved dissemination activities across the entirety of the treated districts, the difficulty of reaching even a majority of citizens in programmed districts proved too challenging. Thus, there is considerable heterogeneity in the extent of programming within districts - some villages received extensive information campaigns on GAS audit results, while others did not. Likewise, some villages received extensive social audit training, other districts received the diffusion of social audit results, and yet others received neither.

The heterogeneity of the spatial concentration of information dissemination and social audit training has important implications for our capacity to detect program effects. This is particularly likely to be the case at the household level. Most importantly, the randomly selected baseline EAs where household interviews took place were likely to have received relatively little direct information dissemination. As discussed in greater detail below, the IE team addressed this challenge by adding an additional EA in the endline sampling. This EA was selected based on information the IE team received regarding the actual implementation of programming. More specifically, the additional endline EA was chosen from where we knew that programming took place.

Third and finally, two rounds of elections - one at the district level and the other at the national level - complicate our capacity to identify program effects. The first elections - a set of DA elections, took place in October of 2015 after having been originally scheduled for the previous year. More than half of political officials in the baseline sample of GAS audit districts were newly elected in October 2015. These newly elected officials were not present for the GAS audits or the exit meetings regarding the audits. Though these elected officials might have experienced subsequent citizen mobilizations as a result of GSAM and did attend DA dissemination events, their more limited exposure to programming might attenuate program effectiveness.

Additionally, the delay in social accountability programming resulted in it running up against the second set of elections - the December, 2016 presidential election. The 2016 presidential election resulted in the loss of the incumbent president John Dramani Mahama to challenger Nana Akufo-Addo, marking the first time in the history of Ghana that the incumbent president sought a second elected term without success. The resulting change in the party of the president has important implications for the composition of district governments, since the president appoints the key DA political official - the District Chief Executive (DCE) - and a third of the members of the District Assembly. The significant turnover in DA officials was in the process of taking place throughout the duration of endline data collection and might have impacted the willingness and responses of outgoing officials to respond to questions about DA budgeting and development projects and also affected the willingness of incumbent officials to provide responses on controversial district issues most relevant to the intervention’s outcomes - such as issues in district budgeting and contracting.

These considerations raise several cautions:
1. The social audit programming took longer to begin and given the longer timeframe for it to take effect, it seems likely that its effects will be more modest at this point.

2. The uneven geographic extent of programming within districts limits our capacity to detect program effects on citizens. The strongest program effects are likely to be found amongst the additional EAs where the IE team is confident that programming took effect.

3. The extensive turnover in DA political officials as a result of first DA elections and then the recent presidential elections might limit our capacity to detect program effects amongst new officials who have only experienced GSAM programming indirectly via citizen action.

4. The original GSAM design was only moderately powered (see baseline report); this generally limits our capacity to detect heterogeneous treatment effects (HTEs) of programming, regardless of whether those HTEs result from regional differences, turnover among DA officials or variation in the intensity of programming across towns and villages.

2.4 Hypotheses
The assessment of the interventions will test a series of hypotheses, which we organize by outcome of interest and target population. We discuss the operationalization of these hypotheses below as well as our statistical approach to multiple hypothesis testing.

Outcome 1: Capital projects and service delivery
The interventions will increase *citizen* access to development projects and the corresponding services (H1), reduce obstacles to completing projects (H2), and increase citizen satisfaction with services (H3).

Within *district governments* (i.e. among administrators and political officials), the interventions will improve district planning (H4), contracting (H5) and oversight of development projects (H6).

Outcome 2: Governance quality
The interventions will improve perceptions of the relative performance of district governments (H7), reduce the incidence of corruption (H8), and increase government transparency (H9).

Outcome 3: District Government Accountability and Citizen Engagement
The interventions will improve the responsiveness of DAs to community needs when it comes to budgeting (H10) and locating development projects (H11).

The interventions will increase accountability for how district officials spend their time (H12) and allocate development projects (H13).

The interventions will increase citizen knowledge about and participation in the budgeting process for development projects (H14).

We discuss how each of these hypotheses are operationalized (Section 4) and tested (Section 5) below. We also discuss in the section on heterogeneous treatment effects that some of these relationships might be conditioned by the quality of the audit reports that the district received. In particular, strong or weak audit reports might condition the relationships posited in H3, H7, H8, H9 and H14.

3.0 RESEARCH DESIGN

3.1 FACTORIAL DESIGN
The study utilizes a randomized control trial design to test the impacts of each intervention of the
program. To assess the individual effect of each intervention, the evaluation design is a three-arm district-clustered, triplet-blocked randomized control trial. Figure 2 illustrates the three treatment arms of the IE, consisting of the different interventions received (GAS Performance Audits with information campaign, CSO Social Audits with information campaign, or pure control). Districts were randomized into these three treatments across Ghana’s 10 regions. A comparison of findings in GAS districts versus control and CSO versus control provides the average program impact on each of these components.

3.2 MATCHING AND RANDOMIZATION PROCEDURE FOR G-SAM

A quantitative matching algorithm was used for the random assignment of districts to GAS, SA, and control groups for GSAM. The goal of the matching and randomization procedure is to ensure that districts in the GAS, SA, and control are as similar as possible. Increasing the similarity of districts prior to programming offers several benefits. First, it decreases the risk that district level characteristics other than the randomly-assigned experimental conditions are responsible for any differences in outcomes between each of the three arms of the program. Differences between the treatment and control districts that are correlated with the program outcomes can lead to biased, incorrect inferences about treatment effects. Second, it increases the robustness of subsequent analyses by decreasing the prevalence of extrapolation across observed variables. Third, directly comparing districts that are most similar on pre-treatment district-level characteristics that predict the outcomes of interest can decrease noise in the data and thereby leading to improvements in statistical power and thus our ability to detect effects of the programming.

In order to ensure that the districts were as similar as possible, we conduct two procedures. First, we create triplicate matches of districts that are as similar as possible. Second, once we had all of the districts matched into similar triplicates, we randomly placed one district from each triplicate into the audit, CSO and control groups respectively. The figure below provides a graphic representation of the matching and randomization procedure, which we further detail.

Figure 2: Evaluation Design

When pre-programming data are available, matching before the randomization procedure is a standard approach to ensure that the study units (districts, in this case) are as similar as possible on key characteristics that are likely to have an impact on program outcomes. Absent initial matching, randomization could produce very different groups of districts in the audit, CSO and control groups simply as a result of bad luck. This potential problem is particularly stark with sample sizes below 300. In such cases, pre-matching units (districts) before randomization ensures that the treatment and control

---

9 See Bruhn and McKenzie (2008) and Greevy et al. (2004).
groups are as comparable as possible, thereby increasing estimation efficiency and statistical power.

In conducting the matches, the goal was to produce sets of districts that are as similar as possible on key characteristics. Doing so required that we identify characteristics most likely to impact the quality of capital projects and political accountability across districts, i.e. the outcomes GSAM aims to improve. We selected six background characteristics for the matching procedure that we expected to be strongly correlated with the outcomes of interest. In particular, we began the matching procedure by first using exact matching on geographic region (north vs. south) and whether a district is new (i.e. created from previous districts during Ghana’s 2012 redistricting) or not. Following this exact matching, we selected a set of additional indicators most likely to predict the outcomes of interest. These included a district poverty index, the ethnic fragmentation of the district, the electoral competitiveness of the district, the district’s most recent (2011) Functional and Organizational Assessment Tool (FOAT) score, and whether the district had previously experienced a GAS performance audit.10

Having identified these key pre-treatment characteristics, we deployed a statistical algorithm to match the districts on the characteristics. There are several approaches to statistical matching and the GSAM program’s design adds a layer of complexity to standard matching procedures. Most matching algorithms are developed to produce matched pairs of units, but GSAM had three arms: pure control, GAS audit, and social audit. In light of this consideration, we aimed to produce matches triplets of units. We used a modified version of Robert Greevy and Cole Beck’s R code for an optimal nonbipartite algorithm to create matched triplets. The algorithm performed better than a standard greedy algorithm and ensured better balance across our treatment and control groups than we could achieve without matching. Further details regarding the implementation of the matching algorithm are available in the GSAM baseline report.

Once the triplets of districts were identified, the randomization procedure was straightforward. One district of each triplet was randomly placed in the GAS audit, social audit and pure control groups as determined by a random number generator. Within each triplet, the district with the highest number was assigned to the control group. The district with the middle number was assigned the social audit programming and the district with the lowest number was assigned to the GAS audit programming.

The baseline report on the G-SAM IE provides evidence that matching districts on pretreatment characteristics contributed to balance across control and treatment districts both in the characteristics of respondents and in survey responses. The midline data also shows good balance. In short, given data constraints the matching and randomization procedures have done a good job of generating balance between treatment and control districts.

It is important to note that although Ghana is comprised of 216 districts as of 2017, our district sample consists of 150 districts. 52 of the 216 districts are classified as “metropolitan” or “municipal”, meaning they are more urban. The programming was not aimed at these more urban districts. Of the remaining universe of 164 traditional districts, the 150 districts for GSAM were selected on the basis of the best matches as described above. Our inferences concern the 150 districts in our sample rather than the broader population of all districts in Ghana. Although this may limit the generalizability of our results to other contexts, matching increases confidence in the unbiasedness of our estimates and should improve the precision of our estimates for a narrower population of districts.

10 These performance audits were done in ten MMDAs in conjunction with DANIDA.
3.3 SAMPLING
There were three stages of sampling - District, Enumeration Area (EA), and household. The district sampling process is described above while the enumeration area and household sampling is described below.

Household Enumeration Areas
At the baseline, the IE team selected the 150 districts and 50 matched triplets according to the criteria described in the matching procedure section above. At the second stage, two enumeration areas (the primary sampling units for the survey) were selected from each of the 150 districts into our sample. Using the 2010 Ghana Population and Housing Census Enumeration Areas (EAs) List as a sampling frame, the EAs in each of the selected districts were categorized into two strata (rural and urban). One rural EA and one urban EA were randomly selected from each district, subject to two constraints on the randomization. The randomization structure was constrained to prevent the selection of neighboring EAs and those characterized by the Ghana Statistical Service (GSS) as “remote”, i.e. unreachable by car or motorcycle. Overall, the proposed household sample size for the original RCT evaluation design was 3000 - 150 districts with 10 households sampled from each of the two EAs.

Deviations from planned intervention implementation (see section “Intervention Timing, Implementation Challenges and Cautions” above), required the IE team to alter the original IE plan. The GAS audits and the corresponding dissemination of citizen scorecards occurred in many villages that did not fall within the baseline survey enumeration areas, and their dissemination did not reach the entirety of each of the GAS audit districts. Likewise, the social audit scorecards were delayed in distribution and did not reach residents across the entirety of each of the SA districts. In other words, although programming took place in all districts according to plan, it did not necessarily take place in all of the planned places in those districts.

Two possibilities arise from the fact that the programming and (particularly) citizen outreach were less extensive than originally planned. First, the baseline EAs may have never received GSAM programming. Second, any attempt to identify the effect of programming should necessarily investigate those places where programming did, in fact, take place. Thus, if we rely only on the baseline enumeration area sampling for the endline impact evaluation, the resulting analysis will identify the effect of the intent to treat. However, the substantively more relevant quantity of interest in such an evaluation is the actual effect of the treatment. In cases of perfect compliance with the programming and where the programming had taken place across the entirety of each district where it was planned, estimates of the effects of the intent to treat are identical to the effect of the treatment. However, a weak relationship between the intent to treat in a given enumeration area and exposure to the treatment would attenuate our estimates of the effects of the actual program toward zero and potentially lead to significant understatements of the effect of programming on the outcomes of interest. Given our knowledge of the actual implementation of the programming, restricting analysis only to the baseline enumeration areas is likely to understate the effect of the treatment.

In order to address these challenges, the evaluation team added one additional enumeration area - comprised of 10 households - to the endline data collection in communities where the maximum suite of programming has taken place. In GAS audit districts, the additional EA was chosen from amongst those towns and villages where scorecard dissemination events had taken place. In the social audit districts, the additional EA was chosen from those villages where CARE and NGO partners had trained local citizens in social auditing and begun the process of diffusing scorecards. As a result of this additional EA in each of the treatment districts, the household survey sample increased from approximately 3,000 at baseline to approximately 4,000 respondents at endline. As enumeration areas surveyed at baseline were selected prior to knowing where high exposure areas would be, these changes better align data collection with the
programming as it was actually implemented. We discuss our approach to estimating program effects across the additional and original EAs in Section 6 below.

**Administrative and Elected Official Surveys**

The administrator and politician surveys aimed to conduct structured interviews 750 DA administrators (5 per district) and 750 local political officials (5 per district) across the 150 study districts for the endline data collection. This endline target sample size mirrors that at baseline.\(^{11}\) Due to variation in availability and accessibility of officials, the politician and administrative official surveys relied on systematically-obtained convenience samples of those most knowledgeable about district development projects and budgeting. In sampling officials for the political and administrative official surveys, enumerators were told to prioritize officials involved in the development and management of the district’s development project budget. To facilitate consistency of sampling, the IE team provided the enumerators with a list of official roles and the prioritization of those officials. The list of Political Officials, from highest priority to lowest priority, included:

- District Chief Executive
- Presiding Member (PM) of the DA
- Chair of the Works Sub-committee
- Chair Development Planning sub-committee
- Chair of the Finance Sub-committee
- Chair of the District Executive Committee (if different from DCE or PM of the DA)
- Chair of the Social Services Sub-committee

The Administrative Officials, from highest priority to lowest priority, included:

- MMDA Internal Auditor
- District Budget Analyst/Budget Officer Development Planning Coordinating Council
- District Coordinating Director (i.e. the head of the DA administration)
- Chair of District Tender Board (Note: only if it is not the DCE)
- District Finance Officer/ Comptroller
- District Planner/Director of the Physical Planning Department
- MMDA Planning Officer (if there is one)
- District Works Engineer
- Assistant District Engineer
- District Procurement Officer

This list of administrative officials is somewhat different from the list provided during the baseline survey. As the programming staff and the IE team learned more about district budgeting for development projects, we agreed to replace the Director of Social Welfare/Community Development and the Director of Education, Youth and Sports Development with the District Planner/Director of the Physical Planning Department and the District Procurement Officer. We also added the District Works Engineer and

\(^{11}\) A limited midline survey was conducted in April and May of 2016 among administrators and political officials. The survey did not take place in the 50 social audit districts because the audits were still in progress, DA dissemination events had not occurred, and the citizen outreach component of the SA programming had not yet occurred. Thus, the midline report was based on a smaller survey of slightly less than 1,000 DA officials from the control and GAS districts. The midline survey instruments sought to assess whether overall levels of service delivery, corruption, governance effort, budget procedures, and other indicators of local government performance varied between control districts and the GAS-audited districts.
Assistant Engineer, who play an important oversight role during the construction of development projects. Enumerators and field supervisors were instructed to use these lists until they reached the target sample size of 10 officials for each district. Enumerators and field supervisors were also instructed to aim for balance between Political and Administrative officials in each district. District officials were contacted in advance of the survey team’s arrival in the district in an effort to establish meetings with those higher on the prioritization list.

3.3 Endline Challenges
In summary, there were several challenges to the endline data collection and evaluation:

1. The challenges in scaling the community resources necessary to conduct the SA programming resulted in a delay in that portion of the programming. While all of the requisite social audits were conducted before endline, there were delays in disseminating district-wide SA scorecards. As described above, several of the dissemination activities were less widespread than initially envisioned and/or ongoing during endline data collection. These factors may limit the capacity of the evaluation to uncover program effects in the SA districts.

2. Presidential election, DA budgeting, and resulting DA turnover: The presidential election in November, 2016 consumed much of the energy and attention of the Ghanaian government. The passage of DA budgets was delayed, which complicated the timing whereby citizen pressures might be brought to bear on the development of DA Annual Action Plans. Most significantly, the victory of the NPP candidate provided newly elected President Akuffo-Addo the opportunity to appoint new district chief executives and ⅓ of the members of each DA. Given that these new DCEs and DA members would have no history with GSAM, this threatened the IE team’s capacity to measure the impact of GSAM on political officials. Thus, the IE team had to rush to get the endline into the field, and ISSER--the local data collection partner--had to hire a particularly large survey team of 61 enumerators, in order to reach outgoing DA political officials before the turnover. These efforts were mostly successful, although the pending turnover did make it more difficult to gain interviews with the requisite officials.
4.0 DATA

Baseline, midline and endline data collection use four primary data collection instruments: (1) household survey, (2) administrator survey, (3) politician survey, and (4) Annual Progress Reports issued by district governments, which provide information on development projects and are used to calculate project completion rates.

QUANTITATIVE DATA SOURCES

For baseline and endline data collection, the household, politician and administrator surveys are structured quantitative instruments. The household survey collects data on a number of measures including livelihoods, satisfaction with local governance and service delivery, quality of district capital projects, responsiveness of district officials and citizen satisfaction and engagement with district-level democratic procedures, in addition to basic social, economic, and demographic characteristics. The politician and administrator surveys seek to assess procedures governing development projects, corruption, governance effort, and adherence to budget procedures.

The household, political and administrative surveys were collected using electronic data capture. Baseline, midline and endline data was collected using Open Data Kit. All data was collected by the Institute of Statistical, Social and Economic Research, a research center housed at the University of Ghana, in the local languages of each region in close cooperation with the G-SAM IE team. Training consisted of a two-week training at baseline, led by the G-SAM research team, and a one week training at endline. A pilot activity was conducted for each wave of data collection. Baseline data collection took place from August 25 to October 1, 2014, and endline data was collected during a two-month period from February – March 2017.12

The study made every effort to re-interview baseline household survey respondents at endline for the original 300 baseline EAs. The survey firm was provided with baseline respondent names, addresses, phone numbers, and landmarks to help identify the correct household panel respondent; that data was protected on a protected data network.

Similar to baseline data collection, endline data collection was subject to a series of quality checks. Given the rush to complete data collection in a timely manner, surveys themselves were not audited as they were at baseline. A series of regular checks conducted by the G-SAM research team screened the data for irregularities as data collection progresses.

Finally, we move beyond opinion data by digitizing DA-generated Annual Progress Reports for 2015 (the 2016 reports have been collected and are being digitized; they will be included, time allowing). These reports provide information on the location, type, funding source, and level of completion of district development projects. DAs nearly always submit them in a timely manner because they are a prerequisite for receiving central transfers in the current fiscal year. Since one of the most important sources of waste in DA budgets is incomplete projects (Williams 2016), we calculate the share of projects that are incomplete for each budget year. If GSAM improves administrative and public oversight of projects, it should be reflected in a lower rate of project non-completion.

12 A limited midline sample of administrators and political officials was collected in the 50 GAS audit districts from March – April 2016. However, this data will not be used for the endline analysis, given the limitations in geographic scope.
5.0 INDICATORS

Per the hypotheses outlined in Section 2.4 above, this study examines the effect of the programming on four sets of outcome measures. In addition to the primary outcome indicators, we construct a series of secondary measures that augment our understanding of the program effects. The majority of dependent and independent variables were constructed through administrator, politician and household survey data. In several cases, we rely on multiple related survey items bearing on a core concept to reduce the noise of measurement. In such cases, we follow the standard approach of calculating z-scores (Kling, Liebman and Katz 2007) as summary measures.\textsuperscript{13} Below, we restate the relevant hypotheses and define the relevant measures. Tables 1 (household survey) and 2 (administrator and politician survey) provide a summary of the indicators and details on how we create indices for those models where we rely on multiple measures of the same construct. In a number of cases, we focus specifically on primary (“basic”) schools because these represent far and away the most common DA-built development projects. We also examine a number of measures bearing on development project completion because project abandonment is a major problem in the districts; Williams (2016) reports that about \(\frac{1}{3}\) of started projects are never completed and these incomplete projects account for 20 percent of all capital spending.

Table 1: Household Indicator Table Here
Table 2: Administrator and Politician Indicator Table Here

DEPENDENT VARIABLES

OUTCOME FAMILY 1: CAPITAL PROJECTS AND SERVICE DELIVERY

Outcome Family 1, Capital Projects and Service Delivery, is focused primarily on a series of variables pertaining to how citizens perceive the quality of district capital projects and health and education service delivery outcomes. Indicators from Outcome Family 1 are constructed out of the household survey and will be used to test the following hypotheses:

The interventions will increase citizen benefits from development projects and the corresponding services (H1), reduce perceived obstacles to completing projects (H2), and increase citizen satisfaction with services (H3).

Primary Indicators:

- Index of family and community benefits from recent development projects (H1).
- Ordinal measure of satisfaction with physical condition of school facilities (H1).
- Index of citizen satisfaction with DA building of: a) roads; b) health facilities; and c) schools (H2).
- Ordinal measure of project delays: Perception of the frequency with which work delays happen on development projects (H2).
- Index of citizen satisfaction with services: index of respondent satisfaction with basic education, health services and public toilets (H3).

\textsuperscript{13} These are calculated by: a) recoding all variables such that better outcomes have higher scores; b) standardizing each recoded variable by subtracting the control group mean from the district-level average, and dividing by the control group standard deviation; and c) averaging across the standardized variables.
Within **district administrations**, i.e. among administrators and political officials, the interventions will improve district planning (**H4**), contracting (**H5**) and oversight of development projects (**H6**).

**Primary Indicators (administrator and politician survey):**

- Index of the strength of internal institutional oversight of projects as measured by the existence and influence of a DA Audit Report Implementation Committee (**H4**).
- Continuous measure of the share of development project that are abandoned reported by administrators and political officials (**H4, H5, H6**).
- Continuous measure of the share of development projects completed in 2016, as collected from *Annual Progress Reports* (**H4, H5, H6**).
- Index of the quality of project implementation over the past year as measured by building delays, abandonments, and on-time completions (**H6**).

**OUTCOME FAMILY 2: DISTRICT GOVERNANCE QUALITY**

Outcome Family 2, District Governance, will be assessed through variables measuring the quality of district governance, with special reference to district capital projects. Indicators from Outcome Family 2 will be used to test the following hypotheses:

The interventions will, reduce the incidence of corruption (**H8**), increase perceived government transparency (**H9**), and improve citizen’s perceptions of the relative performance of district governance (**H7**).

**Primary Indicators (citizen surveys):**

- Index of perceived personal enrichment by DA officials, including: a) the DCE; b) elected DA member; and c) DA professional staff (**H8**).
- Index of perceived governance obstacles to development as measured by citizen perceptions that: a) lack of citizen participation; and b) corruption among district officials are obstacles to building better development projects (**H7, H8**).
- Index of relative governance performance as measured by citizen perceptions of DA: a) corruption; and b) inefficient spending, relative to other DAs (**H7**).

**Primary Indicators (administrator and politician surveys)**

We also test hypotheses 7-9 on data collected from district administrators and politicians:

- Index of the extent to which corruption: a) is a problem in the preparation of Annual Action Plans; b) is a problem in the execution of Annual Action Plans; c) in the DA is an obstacle to building development projects that serve the needs of citizens; d) among contractors is an obstacle to building development projects that serve the needs of citizens (**H8**).
- Index of transparency in: a) the planning; and b) the contracting of district development project (**H9**).
- Index of relative governance performance as measured by perceptions of DA: a) corruption; and b) inefficient spending, relative to other DAs (**H7**).

**OUTCOME FAMILY 3: DISTRICT GOVERNMENT ACCOUNTABILITY AND CITIZEN ENGAGEMENT**

Outcome Family 3 is evaluated through variables that measure citizen engagement with district government and the responsiveness of district officials to public pressure. Indicators for Outcome Family 3 will be used to test the following hypotheses:
The interventions will improve the responsiveness of DAs to community needs when it comes to budgeting (H10) and locating development projects (H11).

**Primary Indicators (Administrator and Politician surveys):**

- Indicator variable of the importance of targeting communities with the most needs in the allocation of development projects.
- Ordinal measure of the extent to which the lack of citizen engagement represents an obstacle to better development projects.
- Index of the existence and activism of the DA Public Complaints Committee.

**Primary Indicators (Citizen survey):**

- Ordinal measure of perceived importance of community needs in where the DA locates projects.
- Index of citizen participation: whether community was consulted on the largest recent project in the area, whether the DA or town council hold public meetings to establish development priorities (H11).

The interventions will increase accountability for how district officials spend their time (H12) and allocate development projects (H13).

**Primary Indicators (Administrator and Politician surveys):**

- A continuous measure of the share of work time that administrators and politicians spend responding to citizen concerns and/or working with community/civil society groups.

The interventions will increase citizen knowledge about and participation in the budgeting process for development projects (H14).

**Primary Indicators (Citizen survey):**

- Whether respondents have participated in community meetings aimed at developing priorities for development projects.

In addition to these measures to test our hypotheses, we also assess the extent to which citizens and DA officials in programming committees are aware of G-SAM, whether it impacted their engagement with the DA, and their overall assessment of it. Analysis of these questions serves two purposes: 1) If G-SAM is to have any direct effect, it must be via some awareness of the programming; and 2) As discussed in Section 2 above, there is heterogeneity in the extent of programming across communities within programmed districts and potential spillovers between them. The measures include:

**Citizen Survey:**

- Whether the respondent has heard of G-SAM
- Whether the respondent has attended a G-SAM meeting
- Whether the respondent has contacted a DA official due to G-SAM
- Whether and how the respondent’s involvement in DA politics has changed due to G-SAM
- The respondent’s overall assessment of G-SAM,
- And whether respondents think G-SAM should continue.

**Administrator and Politician Surveys:**
• Whether the respondent has heard of G-SAM
• Whether citizens have mentioned G-SAM to the respondent on others in their office
• Whether other members of the district government have mentioned G-SAM to the respondent on others in their office
• Whether and how the district government has taken any action in response to G-SAM
• Whether and how citizen engagement with district government has changed as a result of G-SAM

Indicator Tables 1 and 2 above provide further details on how each of these measures are produced.

INDEPENDENT VARIABLES: TREATMENT INDICATORS

We test for the effect of two different treatments and one form of heterogeneous treatment effects. They are measured by the following indicators:

• An indicator variable that equals 1 if a district was assigned to receive the GAS audit intervention, 0 if a district was not assigned to receive the GAS audit intervention.
• An indicator variable that equals 1 if a district was assigned to receive the Social Audit intervention, 0 if a district was not assigned to receive the Social Audit intervention.
• An indicator variable that equals 1 for “new” enumeration areas in treatment districts, which indicates high-intensity treatment EAs, a key potential source of treatment heterogeneity. This is only produced for household analyses.

INDEPENDENT VARIABLES: CONTROL VARIABLES

As described in Section 6 below, some models include a set of control variables. In the household analysis, these include:

• A household asset index, calculated as the first component in a principal components analysis conducted on a series of asset questions.\(^\text{14}\) As reported in the GSAM baseline report, this is the one variable on which there was not pretreatment balance.
• Key respondent characteristics, including gender, age and ethnicity
• District-level NDC presidential vote share in 2012.

The household- and respondent-level measures account for standard features of respondents likely correlated with attitudes toward and experiences with the public sector. The district-level vote share measure addresses the potential for the central government’s fiscal and development project resources to be targeted toward the president’s party.

In the analysis of the Administrator and Politician data, we introduce a measure for ethnicity and the length of time respondents have been in office. This latter addresses the potential for: a) newly rotated administrators into treatment districts who have experienced little programming; or b) for newly rotated administrators into control districts who were previously in treatment districts, who might have extensive experience with GSAM.\(^\text{15}\) See the discussion of potential spillover effects in Section 6 below.

\(^{14}\) The asset questions include whether or not the household has a stove, refrigerator, electric fan, radio, car, bicycle, cell phone, tv, cattle, a sewing machine, and a scooter.

\(^{15}\) As noted above, there was considerable turnover in DA elected officials courtesy of the 2015 election, and administrators are regularly rotated across districts. These factors could reduce the impact of GSAM programming to the extent officials are recently arrived to their positions. It is theoretically possible that administrative rotation occurred as a result of GSAM activities and is, therefore, post-treatment, but qualitative evidence from GAS and
INDEPENDENT VARIABLES: DESIGN EFFECTS
To control for the method of randomization (see Section 6 below), we introduce matched triplicate fixed effects.

6.0 Estimation Strategy

PRIMARY ANALYSIS I: MAIN EFFECTS FOR CROSS-SECTIONAL ENDLINE SAMPLE
The IE is designed to rigorously assess the direct impacts of the G-SAM interventions on the three outcome families described above. The analysis will test the impact of G-SAM on the outcomes described in Section 5 (“Indicators”) at the household and community level. Our estimation will compare the following two groups:

**Group 1)** Each treatment arm (GAS audits, social audits) to the pure control.

**Group 2)** GAS audits to the CSO-led social audit campaign.

We present two sets of results. The first, most straightforward approach reports the intention-to-treat effect (ITT) as the difference in mean outcomes between treatment and control districts without “control variables” and without accounting for any variation in the intensity of treatment within districts (i.e. the third, new EAs). There is debate as to the costs and benefits of introducing additional pretreatment measures as covariates. Since such an approach can improve the precision of estimated program effects and thereby reduce the prospect for false negatives, we also report results with a set of baseline covariates.

We use OLS to estimate the effect of each treatment arm of interest using the following baseline specification:

\[
Y_{ijd} = \beta_0 + \beta_1 T_{ij}^{GAS} + \beta_2 T_{ij}^{SA} + \phi_d + u_{ij}
\]

where \(Y_{ijd}\) is the outcome measure of household \(i\) in district \(j\) in triplicate \(d\). As described in Section 5, most outcome indicators are constructed from survey data and are continuous, dichotomous and ordinal. We also include matched triplicate fixed effects, \(\phi_d\) per Bruhn and McKenzie (2009) to account for design effects.\(^{16}\) We rely on OLS (rather than a combination of linear and non-linear models) across all these variable types in light of evidence on the robustness of OLS in experimental settings\(^ {17}\). \(T_{ij}\) is the treatment dummy for each of the two treatment arms of interest, and \(u_{ij}\) are robust standard errors clustered at the district level using Huber-White sandwiched standard errors (Lin et al., 2013).

In a second set of models we add a set of pre-treatment covariates \(X_{ij}\), described in Section 5 above. Thus, we estimate:

\[
Y_{ijd} = \beta_0 + \beta_1 T_{ij}^{GAS} + \beta_2 T_{ij}^{SA} + \beta_3 X_{ij} + \phi_d + u_{ij}
\]

**PRIMARY ANALYSIS II: HETEROGENEITY IN THE INTENSITY OF TREATMENT**
As discussed in Section 2.3 above, there was considerable heterogeneity in the extent of treatment within districts. While both interventions were initially conceptualized as district-wide, diffusing information

---

\(^{16}\) See Bruhn and McKenzie (2009), who show that failure to control for the method of randomization typically produces overly conservative standard errors.

\(^{17}\) See Judkins and Porter (2016).
across all, or even most, district citizens proved too challenging. This heterogeneity is most marked in the extent of information dissemination across communities within GAS and SA districts, where many communities received no dissemination events. The use of radio jingles and billboards, moreover, makes it impossible for the IE team to know exactly how much programming each community had experienced. As described in Section 3.3 above, the IE team addressed this ambiguity by supplementing the baseline household enumeration areas with an additional one where we did know that programming had taken place; these EAs were chosen from a list of communities that implementers targeted with a full range of community dissemination. Thus, in a second set of analysis we include an indicator variable for these third EAs, which we interact with the treatment dummies to estimate:

\[ Y_{ijd} = \beta_0 + \beta_1 T_{ij}^{\text{GAS}} + \beta_2 T_{ij}^{\text{SA}} + \beta_3 E_j + \beta_4 E_j \times T_{ij}^{\text{GAS}} + \beta_5 E_j \times T_{ij}^{\text{SA}} + \varphi_{d} + u_{ij} \]

\[ Y_{ijd} = \beta_0 + \beta_1 T_{ij}^{\text{GAS}} + \beta_2 T_{ij}^{\text{SA}} + \beta_3 E_j + \beta_4 E_j \times T_{ij}^{\text{GAS}} + \beta_5 E_j \times T_{ij}^{\text{SA}} + \beta_6 X_{ij} + \varphi_{d} + u_{ij} \]

where equation 2a excludes the control variables, \(X_{ij}\) described above, and equation 2b includes them.

**MULTIPLE TESTING CORRECTION**

Given the large number of hypotheses that we will test in our statistical analyses and our reliance on null hypothesis significance testing, we expect our inferences to lead to a non-negligible count of Type I errors (“false positives”) and Type II errors (“false negatives”). Our evaluation will report both raw \(p\)-values that are unadjusted for the multiple comparisons we make throughout our analysis and also more conservative \(p\)-values adjusted for multiple comparisons using the Benjamini & Hochberg (1995) False Discovery Rate Correction. Our main findings and summary sections will rely on the uncorrected values, because we are analyzing a number of closely related interdependent outcomes and, therefore, the standard corrections for the false discovery rate are likely too conservative (Gelman, Hill, and Yajima, 2012). Nevertheless, we will clearly note in the body of the main report when Benjamini and Hochberg-adjusted \(p\)-values are substantively different and suggest a conservative interpretation.

**SECONDARY ANALYSIS 1: MAIN EFFECTS FOR PANEL DATA**

The secondary analysis for the evaluation focuses on a panel analysis of household and administrator data. Equations [3a] and [3b] below will be applied to the analysis of the panel data:

\[ Y_{ijd} = \beta_0 + \beta_1 T_{ij}^{\text{GAS}} + \beta_2 T_{ij}^{\text{SA}} + Y_{ij}^{0} + \phi_{d} + u_{ij} \]

\[ Y_{ijd} = \beta_0 + \beta_1 T_{ij}^{\text{GAS}} + \beta_2 T_{ij}^{\text{SA}} + \beta_3 X_{ij} + Y_{ij}^{0} + \phi_{d} + u_{ij} \]

where \(Y_{ijd}\) is the outcome measure of household or administrator \(i\) in district \(j\). \(T_{ij}\) is the treatment dummy for each of the two treatment groups of interest. \(Y_{ij}^{0}\) is the baseline vectors for the outcome measure, \(\phi_{d}\) is the matched triplicate fixed effects, and \(u_{ij}\) are robust standard errors clustered at the district level, using Huber-White sandwiched standard errors (Lin et al., 2013). Equation [3b] also includes \(X_{ij}\), a vector of control variables, described in Section 5, including included imbalanced covariates.

Panel analysis is a secondary technique for the study due to two main reasons. First and foremost, there is the programming heterogeneity problem that has been described extensively in Sections 2, 3 and 5 above. 

---

18 Gelman and his co-authors note here that for most social science studies, where the effects may be small but are unlikely to be exactly zero, the corrections are likely too conservative.
The implication of this heterogeneity is that the greatest treatment impact will most likely be found in the third EA that was added at endline and for which there is no equivalent baseline measures. Second, due to changes in the program design since baseline, updates were made to the endline instruments to better capture program effects. This restricts the number of indicators that can be used for panel analysis because of differences across the baseline and endline questionnaires.

Due to panel attrition between baseline and endline, the household analysis will explore outcomes across three subsets of the data:

- A respondent panel, including only exact matched respondents between the two survey waves;
- A household panel, including cross-sectional matches for households between the two survey waves;
- A combined household and respondent panel, including all observations for the two subsets noted above.

Similarly, given attrition for the administrator data between baseline and endline, the analysis of administrator data will present findings for two subsets of data (1) the respondent panel, including only exact matched respondents between the two survey waves; and (2) a district panel, including the exact matched respondents and cross-sectional matches for administrators based on job position. Given the small sample size of the administrator sample, we expect this analysis to be underpowered to detect small or moderate treatment effects.

SECONDARY ANALYSIS II: HETEROGENOUS TREATMENT EFFECTS

As discussed above, the primary source of HTEs is variation in the extent of community-level programming. A second potential source of heterogeneity is the quality of the district audit scores that were reported to citizens and district officials via scorecards. The interventions might improve citizen’s perceptions of district governments, reduce perceived corruption, and increase perceived transparency when district scores on GAS or social audits are positive; the interventions are likely to decrease the outcomes above when district scores on GAS or social audits are negative. More specifically, it could be that the interventions will:

- Improve citizen satisfaction with services (H3) when district scores on GAS or social audits are positive and decrease the outcomes above when district scores on GAS or social audits are negative.
- Improve citizen’s perceptions of the relative performance of district governments (H7), reduce perceived corruption (H8), and increase perceived transparency (H9) when district scores on GAS or social audits are positive, and decrease the outcomes above when district scores on GAS or social audits are negative.
- The interventions will improve citizen participation in the budgeting process (H14) when district scores on GAS or social audits are positive, and decrease participation when district scores on GAS or social audits are negative.

In secondary analysis, we test for these potential heterogeneities with an indicator variable, which takes on a value of 1 when district scores are above the median and 0 otherwise. Since audit scores are not available for control districts, we cannot interact them with treatment status. Instead we estimate the conditional average treatment effect (CATE) per equations 1a and 1b separately above and below the mean audit score.

---

19 The GAS and social audit scores are not directly comparable, so we rank districts by score within each treatment arm and define “good” scores as those above the average.
SECONDARY ANALYSIS III: SPILLOVERS

There are two potential sources of spillovers from treated to control districts. First, the local press has become increasingly aware of G-SAM in treatment districts, and the programming itself has involved hundreds of radio jingles to diffuse GAS audit results and promote citizen information and participation. To the extent that radio or print media transcend district boundaries, this might serve as a vector for the transmission of information from treatment to control districts. It is unclear, however, what the nature of the spillovers would be. On one hand, citizens in control districts might become sensitized to opportunities and problems bearing on development projects in their own districts; this could foster activism, improve project outcomes, and attenuate our capacity to identify program effects. On the other hand, citizens in control districts might ignore G-SAM-related news that bears on neighboring districts, in which case spillovers would be minimal to none. Given that we are unable to assess the extent to which media reporting has spilled across treatment and control districts, we take a second best approach, namely in secondary analyses we assess the robustness of our results to the distance of household respondents from the nearest treatment district centroid. To the extent local media diffuse news across district boundaries, the spillovers should decline in distance.

Second, district administrators are regularly rotated from one district to another. The reasons for these rotations are not well understood, but they do make it likely that: a) some administrators who were subject to G-SAM programming in GAS or SA districts were subsequently placed in control districts; and/or b) some administrators who spent much of the period of programming in control districts were rotated into GAS or SA districts and experienced relatively little programming. Both of these sources of spillovers militate against finding program effects—in the first case because administrators who G-SAM has sensitized to the need for improved budgeting and oversight of development project might take improved practices to control districts; in the second case, administrators who at endline are recorded as being in treatment districts might, in fact, have had little time under G-SAM programming. Absent a registry of all administrative rotations, we rely on a measure of the length of time an administrative respondent has been in their current position as a control variable in the primary analysis.

SECONDARY INDICATOR VARIABLES

The richness of our survey data allows for the creation of a number of alternative dependent variables for each of the outcome families described above. We describe these variables below; secondary analyses will be conducted on these outcome measures per Section 6 above.

OUTCOME FAMILY 1: CAPITAL PROJECTS AND SERVICE DELIVERY

Household data:

- Index of citizen access to services: how long it takes members of the household to walk to the nearest health facility, public toilet, and basic (elementary) school.
- Index of citizen satisfaction with the DA in providing: a) water; b) sanitation; and c) maintaining local market places.
- Continuous measure of the length of time it took to build the most recent school.
- Continuous measure of perceived importance of incomplete development projects.

Politician and Administrator Data:

- Ordinal measure of whether political considerations impact the choice of contractors (H5).
- A count of the number problems identified by respondents in the DAs systems for overseeing construction of development projects.
- Ordinal measure of the extent to which respondents agree that DA members gain votes by starting
projects even if they aren’t finished.

- Index of the extent to which respondents think members of the DA would rather complete a smaller number of projects rather than starting a larger number of them but leaving them incomplete.
- Index of deviations in contracting standards as measured by the frequency with which: a) fewer than the required number of contractors submitted bids; and b) large projects were broken into smaller parts to avoid legal requirements for competitive bids.

**OUTCOME FAMILY 2: DISTRICT GOVERNANCE QUALITY**

**Household data:**

- Survey experimental evidence on the share of a 100,000 cedi DA windfall that would be wasted by the district government.
- Index of development-oriented political promises measured as the incidence of community development versus clientelistic benefits promised by politicians.
- Index of partisan favoritism in district governance as measured by citizen perceptions that: a) the DCE; and b) their DA representative will help members of their own party at the expense of the people of the district.
- Index of ethnic favoritism in district governance as measured by citizen perceptions that: a) the DCE; and b) their DA representative will help members of their own ethnic group at the expense of the people of the district.
- Ordinal measure of the importance of the relationship between local and national political parties to get development projects in respondents’ community.
- Ordinal measure of the importance of being a member of the ruling party if someone in the respondent’s village wants to get development projects built in their community.
- Ordinal measure of project fairness: Perception of inequity of project provision across communities in the district.

**Politician and Administrator Data:**

- The number of important changes that were made to the DA budget after it was passed.

**OUTCOME FAMILY 3: CITIZEN PARTICIPATION AND DISTRICT ACCOUNTABILITY**

**Household Data:**

- Index of the extent to which respondent holds DA, their elected Assembly member, and the DCE responsible for the quality of local school buildings.
- Indicator variable of whether respondent would turn to a member of the district government for help with their first development priority.
- Ordinal measure of satisfaction with the DCE.
- Ordinal measure of satisfaction with respondent’s elected DA member.
- Index of whether respondent has filed a formal complaint with their DA member or the DA’s Public Relations Complaint Committee.

**Politician and Administrator Data:**

- Index of the extent to which lack of responsiveness, unethical behavior, and corruption contribute to DA members losing election.
SUBGROUP ANALYSIS

USAID has also stated the desire to for subgroup analysis on gender and region, and particularly with respect to whether the program had different effects in the North. In a separate subgroup analysis, we will interact a gender indicator variable with the treatment indicator variables in a model similar to 2a above, but where the third EA dummy is replaced with gender. We anticipate little ability to identify any differences between the North and the rest of the country due to limited statistical power.
7.0 POWER, BALANCE, and MISSINGNESS

In this section we briefly review balance and power calculations as reported in the baseline report. The main points of emphasis are three-fold. First, despite covering nearly every rural (i.e. non-metropolitan and non-municipal) district government in Ghana, GSAM has to generate fairly large program effects for the IE team to be able to detect those effects. Despite sampling large numbers of political officials, administrative officials, and households during data collection, the fact that assignment to experimental conditions occurs at the level of the district limits the precision of our estimates since it is unreasonable to assume that the responses of individuals within districts to the treatment (or lack thereof) are independent of each other. The challenge of detecting moderately sized and small treatment effects becomes more difficult as the heterogeneity of the impact of the programming on the outcomes of interest increases.

This is a challenge inherent to any project that attempts to scale above the village level. In order to accommodate the challenge of identifying program effects, we report significance levels at the .10, .05 and .01 levels (two-tailed test). Second, although confounding due to imbalance on pre-treatment variables is often a threat to impact evaluations - including many randomized control trials - the use of the matching procedure for sampling of districts from the total population of districts led to a sample of districts balanced across a wide range of potential, observed confounders. Third, given the unpredictable reassignment of administrators across districts and the use of radio to disseminate GAS and SA audit results, there is some potential for spillovers.

POWER

At baseline the IE team estimated the minimum detectable effect size (MDES) under several scenarios. The evaluation team consulted with USAID/Ghana and the DRG Center’s governance advisor about what they would judge to be policy important results for the indicators. According to the policy-makers, 7-9% changes were assessed to be measures of success; these are similar to the range of effects seen in impact evaluations for similar programming elsewhere. Given the MDES of the study for cluster level outcomes, the study will be able to detect changes in the range of approximately 6-9% in key indicators. As shown in the baseline report, the study is powered to determine whether the program was successful in generating the desired treatment effects of 7-9% when it comes to budget management; the study is powered to detect a 10-12% treatment effect on a binary indicator of citizen satisfaction with democratic procedures. Further details of the power analysis are described in the baseline report.

The estimated mid-range MDES of .56 for cluster level outcomes and .25 for survey outcomes will enable the study to detect the changes in outcomes seen in the evaluation of similar development programs that evaluate the effect of audits, including Olken (2007) and Ferraz and Finan (2008). In particular, Olken (2007) finds an 8 to 9% reduction in corruption, as measured by missing expenditures, following an increase in government audits for road-building projects. Correspondingly, Ferraz and Finan (2008) find a 7% reduction in reelection rates of officials with violations for corruptions following the publication of audit results through local radios. Based on USAID programming objectives and treatment size expectations from the academic literature, these detectable cluster and household level effects represent large program "successes". Put differently, the evaluation is not powered to uncover modest program effects.

BALANCE

Although confounding due to imbalance on pre-treatment variables is often a threat to impact evaluations - including many randomized control trials - the use of the matching procedure for sampling of districts from the total population of districts led to a sample of districts balanced across a wide range of potential, observed confounders. As described in the baseline report, the only variable on which there was balance imbalance is an index of household assets. As described above, we introduce this index as a control
TREATMENT OF ATTRITION AND MISSING VALUES

The attrition rate will be calculated from the baseline sample. Several tests will be run after endline data collection is complete in order to determine if there is attrition bias. These tests include t-tests and ANOVA tests to answer the following questions:

- Is the magnitude of attrition different between treatment and control households?
- Are the baseline characteristics of attrited households in the control group significantly different than the baseline characteristics of the attrited households in the treatment group?

If statistical tests reveal attrition to be happening at random, and attrition is rare enough to not strongly affect the power of the design, attrited households will be dropped from the analysis. However, if statistical tests reveal non-random differences in magnitude or baseline characteristics of the attrited in the treatment and control areas, the analysis will adjust the sample through the use of imputation, weighting or Lee bounds so the share of observed individuals is equal for both treatment and control groups.

Missingness is likely to exist in the survey data that underpins our outcome measures. We drop from the analysis questions for which more than 30 percent of responses are missing. When missingness is lower, we impute missing values using the method proposed by Honaker and King (2010); we use religion, level of education, ethnicity, an asset index, region, and questions bearing on how often respondents follow news and have difficulty getting enough food and clean water, NDC vote share in 2012, as well as the outcome measures as predictors. We will consider the sensitivity of imputed results to complete case analysis. We remove variables that have 90 percent of observations with the same value for a response category from the analysis. These variables are not included as covariates or outcome indicators.
8.0 ETHICAL CONSIDERATIONS

Participation in the study was voluntary, and all respondents were required to give their informed consent at the beginning of the survey process. The GSAM IE team received Institutional Review Board approval for the survey instruments from SI on June 20, 2014 and again on December 29, 2016; minor modifications were made to questions for the endline surveys. Informed consent was received from each participant after reading a statement about the purpose of the research, the content of the survey, any risks or benefits, and the time commitment. Participants were assured their participation was voluntary and could be withdrawn at any point and that their answers would be kept confidential.

As described in section 4.0 Data, quantitative data was collected through the ODK platform on Android tablets. Tablets are password protected, and data was uploaded to an encrypted server when network connectivity was available. Data is stored on password encrypted computers, with PII removed. Identified data will be stored on Duke University’s Protected Data Network.
<table>
<thead>
<tr>
<th>Description</th>
<th>Operationalization</th>
<th>Survey Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CAPITAL PROJECTS AND SERVICE DELIVERY</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Benefit from development projects</td>
<td>Z-SCORE (C4, C5)</td>
<td>C4. How much has your FAMILY benefited from the project?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C5. How much has your COMMUNITY benefited from the project?</td>
</tr>
<tr>
<td>Satisfaction with school facilities</td>
<td>C27</td>
<td>C27. Overall, how satisfied are you with the CONDITION OF THE FACILITIES at the BASIC school your child/children attend?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C28. How satisfied are you with the DISTANCE TO the BASIC school that your child/children attends?</td>
</tr>
<tr>
<td>Prevalence of project oversight</td>
<td>VIG2_3</td>
<td>READ: Samuel's district government began the construction of a school a year ago. Construction of the building was very slow, and no one has been seen working on it for six months. VIG2_3. How often do situations like those in Samuel's district happen in your district?</td>
</tr>
<tr>
<td>Satisfaction with building</td>
<td>Z-SCORE (D7, D8, D9)</td>
<td>D7. How satisfied are you with how the District Assembly is BUILDING ROADS, or have you not heard enough to say?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D8. How satisfied are you with how the District Assembly is BUILDING BASIC HEALTH FACILITIES, or have you not heard enough to say?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D9. How satisfied are you with how the District Assembly is handling the BUILDING AND MAINTAINING OF SCHOOLS, or have you not heard enough to say?</td>
</tr>
<tr>
<td>Satisfaction with services</td>
<td>Z-SCORE (D11, D11a, D12, C9)</td>
<td>D11. How satisfied are you with how the District Assembly is PROVIDING WATER, or have you not heard enough to say?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D11a. How satisfied are you with how the District Assembly is PROVIDING SANITATION SERVICES, or have you not heard enough to say?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D12. How satisfied are you with how the District Assembly is MAINTAINING LOCAL MARKET PLACES, or have you not heard enough to say?</td>
</tr>
<tr>
<td><strong>GOVERNANCE QUALITY</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Governance obstacles to development | Z-SCORE (C56, C58) | C56. How big of a problem is LACK OF CITIZEN PARTICIPATION AND OVERSIGHT to building better development projects?  
C58. How big of a problem is CORRUPTION AMONG DISTRICT PUBLIC OFFICIALS to building better development projects?

| Knowledge of district’s relative performance | Sum(Indicator if response to D14a = Don't Know, indicator if response to D14b = Don't Know) | D14a. How common do you think GOVERNMENT CORRUPTION is in your district assembly compared to other district assemblies in Ghana?  
D14b. How common do you think INEFFICIENT DISTRICT GOVERNMENT SPENDING is in your district assembly compared to other district assemblies in Ghana?

| Political elites personal benefiting from projects | Z-SCORE (D28, D29, D30) | D28. Sometimes officials benefit personally and financially from public projects. How much do you think your DCE benefits financially from the district’s development projects, or haven’t you heard enough about them to say?  
D29. How much do you think YOUR ELECTED ASSEMBLY MEMBER benefits financially from the district’s development projects? (or have you not heard enough to say?)  
D30. How much do you think THE STAFF OF YOUR DA benefits financially from the district’s development projects? (or have you not heard enough to say?)

| ACCOUNTABILITY |  |  

| Project located based on community needs | C42 | C42. When you think of development projects in your district (such as schools, health clinics, electrification, and markets) how much do you think the NEEDS OF THE COMMUNITY influence where those development projects are located?  

| Attends development project priority meetings | E2 | E2. Have you ever attended such a meeting?  

| Prevalence of opportunities for public input into development | Z-Score(C3, E1, E3) | C1. Please describe the largest development project that has been worked on in your community in the last two years...C3. Was your community consulted in the development of the project?  
E1. Does your district assembly or town council ever hold public meetings to establish development priorities?  
E3. How much influence do you think the people in this community have over decisions the District
Assembly makes about development projects, such as school buildings, health clinics, irrigation ditches, or roads?

**GSAM AWARENESS**

<table>
<thead>
<tr>
<th>Question</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Has heard of GSAM</td>
<td>G0</td>
</tr>
<tr>
<td>Attended GSAM meeting</td>
<td>G2</td>
</tr>
<tr>
<td>Contacted DA due to GSAM</td>
<td>G5</td>
</tr>
<tr>
<td>how did gsam change your involvement</td>
<td>G6</td>
</tr>
<tr>
<td>overall gsam evaluation</td>
<td>G7</td>
</tr>
<tr>
<td>should gsam continue</td>
<td>G8</td>
</tr>
</tbody>
</table>

G0. Have you heard of the Ghana Strengthening Accountability Mechanisms project? It is also referred to as G-SAM.

G2. Have you attended a community meeting organized by G-SAM?

G5. Have you contacted DA officials because of G-SAM?

G6. How has your involvement in your district's politics changed due to G-SAM?

G7. How would you rate the G-SAM overall?

G8. Would you like G-SAM to continue as a program in your district?

**TABLE 2: ADMINISTRATOR AND POLITICIAN INDICATOR TABLE**

<table>
<thead>
<tr>
<th>Description</th>
<th>Operationalization</th>
<th>Survey Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAPITAL PROJECTS AND SERVICE DELIVERY</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strength of formal DA institutions to oversee projects</td>
<td>Z-SCORE (J1, J2)</td>
<td>J1. Does your DA have an ARIC, that is, an Audit Report Implementation Committee?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>J2. Over the LAST YEAR how much influence would you say the ARIC has had on how the district contracts and oversees projects?</td>
</tr>
<tr>
<td>Frequency of project abandonment</td>
<td>D20 and D21 collapsed to a</td>
<td>D20. In your estimation, what percentage of COMMON FUND PROJECTS are abandoned during construction (i.e. never finished) four years from when they</td>
</tr>
<tr>
<td>Project outcomes in past year</td>
<td>Z-SCORE (F1,F3,F11)</td>
<td></td>
</tr>
<tr>
<td>------------------------------</td>
<td>---------------------</td>
<td></td>
</tr>
<tr>
<td><strong>D21.</strong> In your estimation, what percentage of DONOR FUNDED PROJECTS are abandoned during construction (i.e. never finished) four years from when they began?</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>F1.</strong> How often have DELAYS IN CONSTRUCTION happened in THE LAST YEAR?</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>F3.</strong> How often in THE LAST YEAR have CONTRACTORS ABANDONED PROJECTS for which they have received advance payment?</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>F11.</strong> In your estimation, what percentage of large district capital projects have finished on time in THE LAST YEAR?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Project completion rate | Share of projects completed in 2016; calculated from Annual Progress Reports |

**GOVERNANCE QUALITY**

<table>
<thead>
<tr>
<th>Relative district governance comparison</th>
<th>Z-SCORE (G1,G2)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>G1.</strong> How common do you think GOVERNMENT CORRUPTION is in your district assembly compared to other district assemblies in Ghana?</td>
<td></td>
</tr>
<tr>
<td><strong>G2.</strong> How common do you think INEFFICIENT DISTRICT GOVERNMENT SPENDING is in your district assembly compared to other district assemblies in Ghana?</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Corruption impeding development</th>
<th>Z-SCORE (G3,G4,H8,H9)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>G3.</strong> To what extent would you say that CORRUPTION is a problem in the PREPARATION of the Annual Action Plans?</td>
<td></td>
</tr>
<tr>
<td><strong>G4.</strong> To what extent would you say that CORRUPTION is a problem in the EXECUTION of the Annual Action Plans?</td>
<td></td>
</tr>
</tbody>
</table>

READ: There can be many obstacles to building development projects that serve the needs of district citizens. When you think about the preparation of the next Annual Action Plan to what extent do you agree that the following are likely to represent big obstacles to building better development projects in your district?...**H8.** To what extent do you agree or disagree that CORRUPTION AMONG DISTRICT PUBLIC OFFICIALS will represent a big obstacle? **H9.** To what
extent do you agree or disagree that CORRUPTION AMONG CONTRACTORS will represent a big obstacle?

Project transparency: Z-SCORE (H1,H2)

H1. To what extent would you agree or disagree that district capital projects are PLANNED in a transparent manner?
H2. To what extent would you agree or disagree that district capital projects are CONTRACTED in a transparent manner?

<table>
<thead>
<tr>
<th>ACCOUNTABILITY</th>
</tr>
</thead>
</table>
| **Time spent on public service and receiving public input** | B6d + B6e | [READ ALOUD] In a TYPICAL working week, how many hours do you spend on each of the following tasks? If you spend an hour working on something that belongs to more than one of these categories, please include it in both...B6d. Providing services/responding to citizen concerns  
| **Needs for project location** | Recode D13 such that 5 (Targeting areas with the most need) = 1 and otherwise 0 (including Don't Know/Refused) | D13. What is the SINGLE most important influence on where YOU want development projects to be located?  
| **Would public participation improve development?** | H3 | H3. To what extent do you agree or disagree that CITIZENS' NON-PARTICIPATION IN PLANNING AND OVERSIGHT will represent a big obstacle to building better development projects?  
| **Strength of Public Complaints Committee** | Z-SCORE (I3 and I4) | I3. Does the District Assembly have a Public Relations and Complaints Committee?  
| **Strength of Public Complaints Committee** | Z-SCORE (I3 and I4) | I4. How active is the DA Public Relations Complaints committee?  

<table>
<thead>
<tr>
<th>GSAM AWARENESS</th>
</tr>
</thead>
</table>
| **Heard of GSAM** | J14 | J14. Have you heard of the Ghana Strengthening Accountability Mechanisms project? It is also referred to as G-SAM.  
<p>|</p>
<table>
<thead>
<tr>
<th>Topic</th>
<th>Code</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citizen mentions of GSAM</td>
<td>J16</td>
<td>J16. Have citizens in your district mentioned the G-SAM project to you OR others in your office?</td>
</tr>
<tr>
<td>Government member mentions of GSAM</td>
<td>J19</td>
<td>J19. Have other members of the district government mentioned the G-SAM project to you or others in your office?</td>
</tr>
<tr>
<td>Whether district responded to GSAM</td>
<td>J21</td>
<td>J21. Has the district taken any action in response to G-SAM?</td>
</tr>
<tr>
<td>How district responded to GSAM</td>
<td>J22</td>
<td>J22. What action has the district taken in response to the G-SAM?</td>
</tr>
<tr>
<td>Whether citizen involvement changed due to GSAM</td>
<td>J17</td>
<td>J17. Has citizen involvement in your district's politics changed due to G-SAM?</td>
</tr>
<tr>
<td>How citizen involvement changed due to GSAM</td>
<td>J18</td>
<td>J18. How has citizen involvement changed?</td>
</tr>
</tbody>
</table>