Pre-Analysis Plan: Strengthening Fiscal Contracts Through Digital Town Halls in Freetown, Sierra Leone

Kevin Grieco*
Niccolò Meriggi†
Julian Michel*
Wilson Prichard‡§

July 20, 2022

*PhD Candidate in Political Science, UCLA
†Country Economist, IGC Sierra Leone
‡Associate Professor, University of Toronto
§This project was supported by generous funding from the International Growth Center and the International Center for Tax and Development. This study received IRB approval at UCLA (IRB #20-000380) and in Sierra Leone (approved 3/25/2020; amendment approved 5/28/2020). We thank Graeme Blair, Katherine Casey, Darin Christensen, Cesi Cruz, Adrienne LeBas, Giulia Mascagni, Oyebola Okunogbe, Daniel Posner, Soledad Prillaman, and Jon Weigel for helpful comments. We would also like thank participants at the CaliWEPSs IV workshop for their insightful feedback. We thank Jacques Courbe, Emile Eleveld, Yojin Higashibaba, Xenia Rak, and Michael Rozelle for excellent research assistance. Thanks to our wonderful team of supervisors and enumerators. Of course, this project would not have been possible without the contributions and cooperation of the Freetown City Council.

Contents

1 Introduction **4**

2 Digital Town Halls: Motivation **4**

3 Intervention and Treatment Description **6**
3.1 Digital Town Halls . . . 7
3.2 Service Delivery . . . 9
3.3 Service Delivery Notification Calls . . . 10

4 Theoretical Model and Hypotheses **11**
4.1 Primary and Secondary Hypotheses . . . 11
4.2 Causal Mechanisms . . . 11
4.3 Heterogeneous Effects . . . 11

5 Measurement **12**

6 Study Population and Sampling **14**

7 Treatment Assignment **14**
7.1 Matching . . . 14
7.2 Treatment Assignment and Balance . . . 16

8 Data Analysis **17**
8.1 Intent-to-Treat Effect (ITT) . . . 17
8.2 Complier Average Causal Effect (CACE) . . . 19
8.3 Inference . . . 20
8.4 Heterogeneous Effects . . . 20
8.5 Connection to Analyses in Other Projects . . . 20

9 Unbundling the Intervention **20**
9.1 How Does Service Delivery Condition the Effects of the DTH? . . . 21
9.2 Notification Calls . . . 21

10 Spillovers **22**
10.1 Estimating Geographic Spillover . . . 23
10.2 Detecting Social Spillovers . . . 24

11 Power Simulations **25**
11.1 Power Simulations: ITT Analysis . . . 25
11.2 Power Simulation: Heterogeneous Effects Analysis . . . 26

12 Research Ethics **29**

Appendices **34**

A Outcome Summary Statistics **34**

B Descriptive hypotheses **35**
B.1 Descriptive Analysis 1: Who Participates Actively? . . . 35
B.2 Descriptive analysis 2: Who Sets the Agenda? Who Influences Eventual Choices over Services? . . . 35

C Distance to Closest Study Property **36**

D Covariate Summary Statistics **37**

E Project Pictures **38**

F Number of Units in Spillover Estimation **41**

G Manuals and Instructions **42**
G.1 Timeline 42
G.2 Facilitator Guidelines 43

1 Introduction

Does participation in digital participatory budgeting Town Halls increase tax compliance? A large literature examines whether increased taxation spurs demands for accountability and political representation (Bates and Lien 1985; Tilly 1985; Levi 1988; North and Weingast 1989; Ross 2004; Martin 2014; Prichard 2015; Weigel 2020). Yet, experimental evidence on whether taxpayers in fact comply more when given fora to engage with their political representatives and to directly decide over budgets for service deliveries has remained scarce. Regrettably, the prior literature on the taxation-representation nexus largely remains moot on whether greater political inclusion will enhance tax compliance *once democratic institutions are already established*. As we argue, participatory budgeting might induce quasi-voluntary tax compliance—and thus serve as an entry point for better fiscal contracts—in low-capacity democracies. To test our theory, we rely on a field experiment and administrative compliance data from Freetown, the capital of Sierra Leone. If our expectations are born out, policy makers could adopt participatory budgeting interventions similar to the one we test to increase compliance without incurring the administrative and electoral costs associated with purely enforcement-based strategies.

In this project, we cooperate with the Freetown City Council (FCC) to organize 58 digital participatory budgeting Town Halls (DTH) in 30 out of 48 wards of Freetown, Sierra Leone. The research is situated in the context of a city-wide property tax reform that increased the FCC's property tax potential revenue five-fold and collected revenue three-fold. The DTHs take the form of a moderated WhatsApp group chat to which up to 40 resident property owners with confirmed WhatsApp usage are invited. Their task is to deliberate what services are needed most in their ward, to exchange views over priorities with their representatives, and to eventually decide over how the DTH budget of 15 million leones ($\approx$ USD 1,500) should be invested. The DTH process entails four separate phases: Horizontal deliberation, preference articulation and aggregation, vertical interaction, and decision over preferred service. We hypothesize that participation in DTHs increases property tax compliance through four mechanisms: increased willingness for fiscal exchange, enhanced political efficacy, positive updating about tax system fairness and equity, and increased perception that tax enforcement is more likely.

To identify causal effects of the DTHs, we use a matched-pair design to randomize half of 3,618 property owners into treatment. In addition to being invited to join a DTH, treated units receive a notification call from an FCC representative informing them that the selected project has been completed. We observe property-level tax compliance behavior—our primary outcome of interest—through access to FCC administrative records. To nail down attitudinal mechanisms driving this effect, we collect three rounds of survey data: (i) before the DTH, (ii) after the DTH and before the implementation of the selected service, and (iii) after implementation of the selected service.

Digitization continues to be a defining trend affecting social, economic, and political life in the twenty-first century, for rich and poor countries alike. Therefore, we might expect that political leaders in the Global South will be looking for strategies that harness digitization to draw their citizens into virtuous cycles of fiscal exchange and further legitimate leaders' authority. In this study, we provide causal evidence on how governments can use messenger services to improve state-society relations through virtual participatory budgeting.

2 Digital Town Halls: Motivation

Non-digital Town Halls have become a prominent facilitator of citizen-representative interactions since the late 1980s (Sheely 2015, p.252f.). Such deliberative settings have been shown to:

- improve development outcomes (Gonçalves 2014) (but see Mansuri and Rao 2013)
- increase vote shares of participating parties (López-Moctezuma et al. 2022)
- decrease clientelism (Fujiwara and Wantchekon 2013)

- decrease occurrences of violent events (Collier and Vicente 2014)
- increase the political efficacy of participants (Boulianne 2019)
- allow citizens to become more informed (Esterling et al. 2011)
- allow for updating of policy preferences (Barabas 2004; Farrar et al. 2010; Luskin et al. 2014; Sandefur et al. 2020)

Therefore, it is unsurprising that it recently has been argued that the "creation of deliberative spaces where citizens and political elites participate in meaningful conversations with real policy consequences is central to strengthen the quality of weak democracies" (López-Moctezuma et al. 2022, p.73). Yet, there remains much that we do not know about the impacts of such participatory processes, and about different ways of organizing them. Against that background, we make two central contributions.

First, we focus on the underexplored question of how participatory processes impact tax compliance, and the mechanisms through which those impacts occur. A handful of existing studies have argued that participatory processes can lead to increased tax compliance (Sjoberg et al. 2019; Touchton et al. 2019; Torgler 2005). A weakness of this literature is that it focuses on attitudes toward taxation, rather than tax compliance behavior, or that evidence is generated with observational designs (for example, comparing compliance across jurisdictions that varied in their adoption of participatory processes). Even though fiscal contract approaches fundamentally rest on the idea that participatory institutions will induce tax compliance, we currently lack field experimental evidence linking access to participatory institutions to individual-level, administrative compliance data. This study will thus be the first to provide a robust and disaggregated investigation of the impact of participatory processes on tax compliance.

Second, few studies have examined the potential of *digital* Town Halls to enhance political accountability (for reviews of this literature, see Kies 2010; Friess and Eilders 2015). As digitization is expected to be a defining trend for developing countries in this century, policy makers increasingly have to decide whether to offer participatory processes online. As we argue, even if face-to-face interactions are possible (which cannot be taken for granted, as COVID-19 has forcefully shown the world), digital town halls may offer considerable advantages vis-á-vis their offline analogue.

To begin with, participation can be less costly: If access to WhatsApp already exists, participants only need to invest a modest amount of time and mobile data to enter the DTH. Whereas offline THs enable participation only for a short and fixed time period, DTHs can be accessed for weeks and whenever it is convenient for participants. This flexibility reduces the oft significant opportunity costs of participation (Casey 2018). Intuitively, transportation costs—traditionally a barrier to participation especially in rural settings (Sexton 2017, p.35)—are not incurred. Remarkable improvements in internet activity in developing countries—31 % of Sierra Leoneans in 2018 own a phone with internet access (Afrobarometer 2018)—have led to an explosion in social media usage (21.5% of Sierra Leonean report obtaining news through Facebook or Twitter at least "a few times a week" (Afrobarometer 2018). As our study population is property owners in the capital city, we expect these numbers to be even higher in our setting. In our model of mediated interaction through WhatsApp, participation is less costly for political representatives too: All that is required of them is to read a summary of participant contributions and to respond in a limited number of video and voice messages.

Second, perhaps counter-intuitively, we argue that DTHs hold more deliberative promise: In the Habermasian ideal type of deliberative democracy, participants engage in potentially endless communicative action (an exchange of reasoned arguments) as equals until the best argument prevails (Habermas 1975). In offline THs, attendants regularly find themselves unable to make their views known in front of representatives as time constraints only allow for a limited number of contributions. Statements, especially from members of marginalized groups, are often interrupted by other participants (Parthasarathy et al. 2019). In contrast, DTHs allow all participants to make their views known without running the risk of interference by others. Importantly, DTHs alleviate the constraint

of limited attention spans on successful argumentative reasoning: While it is easy to forget what a participant argued a few minutes ago in an offline TH, participants in WhatsApp can just scroll back. Whereas immediate reactions are required offline to ensure that the conversation stays on topic, DTHs enable participants to first reflect on their statement—in theory for multiple days—before posting it. Therefore, the longer time frame in a DTH should increase the argumentative quality of contributions and facilitate perspective taking (as the need for immediate reactions in offline DTHs precludes taking the time to reflect on where someone else's argument is coming from). Finally, we can avoid face-to-face interactions which in group settings under time constraints lend themselves to emotionalized exchanges (more cues are visible—e.g., body language and facial expressions—which make it harder to focus on the merits of the argument alone). Third, DTHs can alleviate one dimension of the well-known gap in political participation by targeting the relatively young who usually are less likely to participate in conventional forms of political engagement. Yet, it is to be expected that DTHs—just like their offline analogue—display additional participation biases (higher ability and willingness to participate among those able to afford smart phones and internet usage, the more educated and literate, those with higher political efficacy (on self-selection in offline TH participation, see Boulianne 2019; Neblo et al. 2010).

However, there are also potential relative disadvantages to the DTH format: The relative anonymity decreases the (reputational) cost of disruptive behavior as participants can choose how much identifying information they provide through their WhatsApp profile. Furthermore, moderating chats can be costly, constrained by the functionalities provided by WhatsApp (messages can only be deleted by who wrote them) and, if done poorly, runs the risk of altering the conversation. The absence of face-to-face interactions can lead to questioning that one is actually talking to ones' representatives and fellow community members. Fortunately, this is less of a concern here as political representatives have prominently associated themselves with the DTH intervention in public. One may argue that voice- and text-based communication is less rich when other cues cannot be observed (e.g., the eyes as an indicator of the sincerity of the speaker). The mediated interaction between participants and representatives relies on trust in the intermediary that is aggregating the information. Perhaps most crucially, while DTHs reduce participation costs for many, those lacking internet/ WhatsApp access cannot participate. Finally, the brevity of text messages may not be conducive to the articulate elaboration of arguments (Jaidka et al. 2019). However, there are no length limitations in WhatsApp and participants have the option to record voice and video messages as well. Through our endline survey and by capturing all DTH conversations, we can measure many of the aforementioned potential disadvantages how prevalent they were.

3 Intervention and Treatment Description

This research takes place in cooperation with the Freetown City Council (FCC) in a context of a city-wide property tax reform two of us helped lead. The reform served to broaden the tax base—less than 50% of the approximately 120,000 properties had been registered previously in the property cadastre—and to make the tax burden more equitable through the introduction of a more nuanced, consistent and transparent property valuation scheme. The mayor publicly announced that DTHs would be held starting in January of 2021. In her messaging, she emphasized that these DTHs are key to secure citizen participation. She stressed that she intends to institutionalize the THs and that future THs will be assigned 20% of the property tax revenue raised in a given ward (see the Freetown City Council's second year Transform Freetown report, pg. 26)

The digital town halls were part of a broader intervention that contained three components: (i) digital town halls, (ii) service delivery, (iii) notification calls about delivered services. Table 1 summarizes the intervention components received by members of the treatment and control groups.

Intervention Component	Treatment Group	Control Group
Digital Town Hall	X	
Service Delivery	X	X
Notification calls	X	

Table 1: Intervention components

Note that we manipulate participation in the digital town halls and reception of a service delivery notification call across groups, holding constant the delivery of services. This implies that the estimand in our primary analysis is the effect of participating in a digital town hall plus having received a notification call, conditional on services being delivered. In additional analyses (described in section 9) we attempt to untangle how service delivery conditions the effect of the DTH on attitudes and behavior (section 9.1) and outline a strategy for isolating the effect of providing property owners with information about recently implemented services (section 9.2).

3.1 Digital Town Halls

In this study, DTHs take the form of WhatsApp group chats. Participants were assigned to one of 58 chat groups, where the number of participants in each chat ranged from 17 to 37 (the median chat group size was 24). All participants in a given chat group owner property in a same ward. The overarching goal of the DTH is for the group to deliberate and decide over how a budget of 15 million leones (about USD 1,500) should be spent in their ward. The budget allocated to the TH does not come from FCC tax revenue given (1) the severity of the budget constraint the FCC faces and (2) that the expected increase in property tax revenue will be accrued after the DTHs have taken place. For these reasons, the funds to be decided over are taken from the project's research budget. However, this is not communicated to the DTH audience, allowing the Mayor and (FCC) political representatives to fully claim credit for the participatory budgeting opportunity, which further ensured buy-in to enable our research.

We designed the DTHs with the goal of enabling both (direct) citizen-citizen and (indirect) citizen-representative interactions. This is reflected in the DTH's four distinct phases: (i) horizontal deliberation, (ii) preference articulation and aggregation, (iii) vertical interaction, and (iv) decision making over services. Videos from political representatives were shared with DTH participants in two ways: videos were posted directly in the WhatsApp group and were available via a Qualtrics link, also posted in chat groups. DTH facilitators requested that participants only use the chat between 7am and 10pm daily, so as to ensure that a facilitator can be present at all times. Participants are free to choose the form in which they would like to participate (text/ voice/ video messages), but were asked to contribute in Krio or English. Please refer to appendix G for the timeline and facilitator guidelines. We completed a pilot DTH in one ward before scaling the DTHs up to our 30 study wards.

3.1.1 Verification and Informed Consent (September, December 2020)

Property owners were only eligible for the study if they were verified WhatsApp users (see section 6 for full sampling eligibility criteria); 1,637 of 1,809 treatment respondents provided informed consent to join their DTH and we verified 1,459 as actually joining. The DTH commenced after introductory videos by both the mayor and the ward councilor, where both representatives explained the DTH process and goals. Additionally, DTH facilitators introduced themselves to participants (both in the group chat and in one-to-one conversations with participants).[1]

[1]The research team hired a local team to act as facilitators, supervised and managed by project RAs.

3.1.2 Phase 1: Horizontal Deliberation (January 2021)

In the first phase of "horizontal deliberation," participants discuss which services they would like to see improved in their ward. This phase is purely horizontal because participants are told that representatives will not be involved at this stage and will not learn about what was discussed. The facilitators begin the conversation by sharing a menu of preferred services taken from a citywide survey we conducted earlier. Participants are then asked which services, within the budget allocated, they would like to see added to the menu that will constitute the choice set for the eventual DTH vote. Overall, this phase serves to offer a "safe discussion space" for citizens that approximates the conditions under which deliberation is thought to work (i.e., the relative equality of citizens exchanging reasoned views on a topic of shared interest, framed to suggest sociotropic concerns—"which service would improve well-being in your ward the most?").

3.1.3 Phase 2: Preference Articulation and Aggregation (January 2021)

The second phase, "preference articulation and aggregation", is the first in which participant statements will be shared with political representatives. Participants are told that this will take the form of an unbiased and anonymized aggregation of their views—performed by the study team—that is presented to both the mayor and the respective ward councilor. While the discussion is still focused on preferred service delivery, participants now articulate preferences towards their representatives.

3.1.4 Phase 3: Vertical Interaction (February 2021)

In the third phase, "vertical interaction," the councilors respond in separate videos to the comments made by DTH participants. These response videos-—one per councilor—allow the representatives to acknowledge the input received and to position themselves to the demands made.[2] This includes highlighting their preferred services, justifications for their service preferences and explaining past and future delivery goals. The representatives also revealed a list of service projects-—pre-determined by the study team-— that participants vote over in the next phase. Participants are then invited to discuss the reactions received from their representatives. We opted for this mediated interaction between citizens and representatives to (i) avoid elite-domination of the TH process and (ii) make realistic time-demands on representatives.

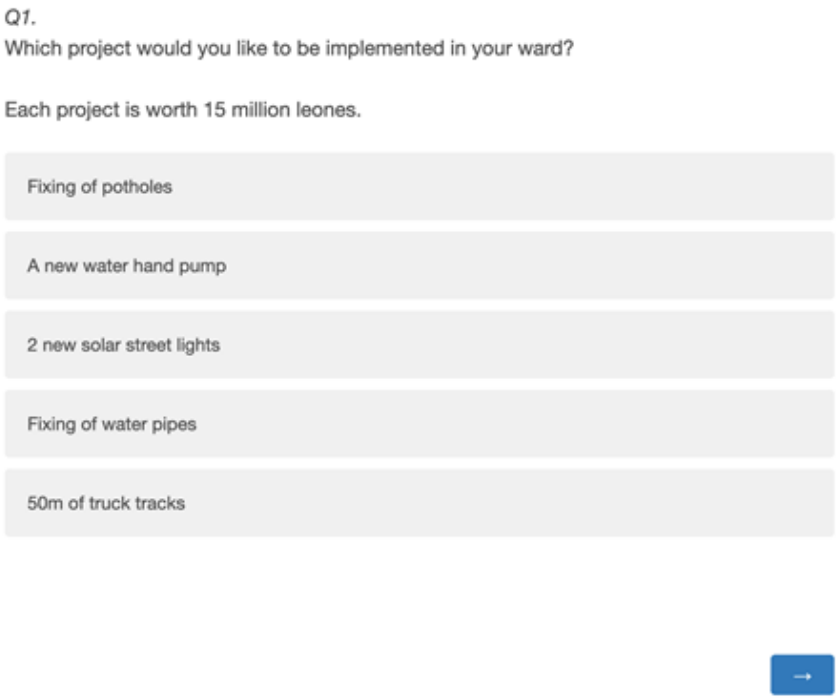
3.1.5 Phase 4: Decision Making (February 2021)

The fourth phase, "decision making," consists of the final vote over how the DTH budget should be spend. This vote could be cast anonymously through a Qualtrics survey. We additionally gave participants the option to inform moderators of their vote in bilateral conversations. The choice set is given by a menu of services—the list provided in phase 1 plus additional items recommended in this first phase that were deemed feasible and within budget by the study team.

The whole budget is dedicated to the service selected by majority rule after aggregating votes to the ward level. The mayor announced the winning project through individualized group messages to each ward. After the announcement, we halted the WhatsApp groups after thanking participants for their contributions. We explained that chats would be used one more time in the future to announce that delivery of the service has concluded. Regarding determining the site of implementation, we explained that we would choose a site in the respective ward at which delivery can be accomplished within budget (as determined by the local implementing partners we cooperated with). When there are multiple such sites, we choose the one which promised the highest utility (because it was most central, closer to higher population density areas, or because of the relative improvement towards the existing local public good in that place).

[2]Where there were two DTHs in a ward, the councilor prepared a video that addressed concerns raised in both DTHs.

Figure 1: Menu of Services



Of a total of 1,637 treatment respondents who consented to be invited to DTH, we confirmed that 1,459 actually joined (89.1%), 921 posted at least one message (56.3%), and 1,016 cast a vote (62%). The median number of messages posted per DTH was 70, about evenly split across text and voice messages.

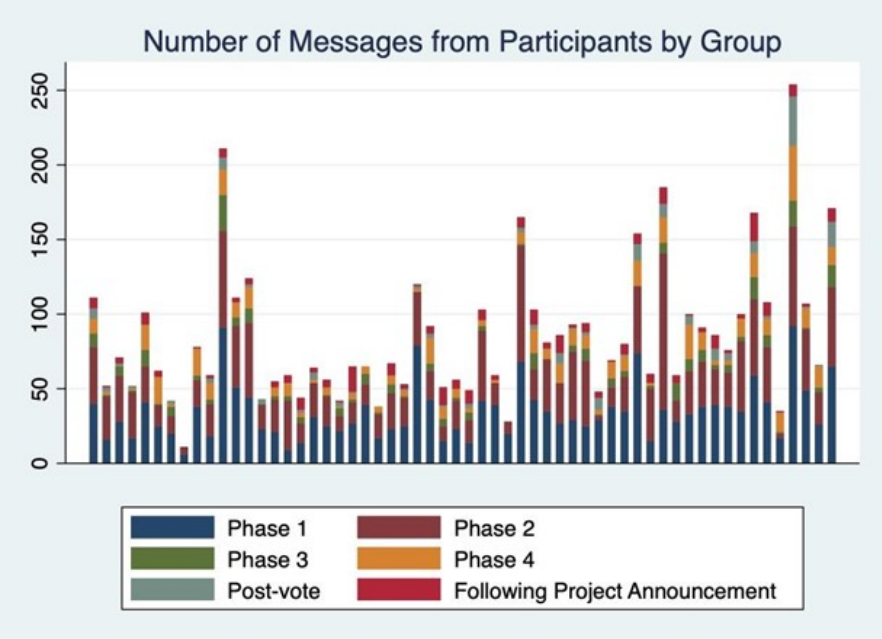
3.2 Service Delivery

Each participating ward received a service project—essentially a local public good both treated and control units in that ward could profit from.[3] Implementation was initially scheduled to start in May 2021, after completion of our midline survey, but was delayed due to negotiations with the delivery firm as well as the underestimated complexity of identifying appropriate delivery sites. Construction began in most wards in October 2021, with the notable exception of one ward (Tengbeh Town) where the FCC needed to provide additional assurances over liability to the delivery company. By the end of 2021 all projects were completed, except the project in Tengbeh Town, which was completed in February of 2022.

Despite multiple rounds of assessment visits of potential delivery sites-—in which project team members were accompanied by an engineer from a construction firm-—there was heightened uncertainty in some wards over whether the selected projects could effectively be delivered there. In consequence, in some wards we had to opt for the delivery of a service that had not been the first choice. In earlier DTH stages, participants were promised to receive the most-voted service *that could be delivered*. This promise was kept in the delivery phase. The only exception occurred when a ward expressed a clear preference for a water-related project (hand pump or fixing of water pipes): When both of those turned out not to be feasible—despite earlier assessments and promises by the construction firm—we opted to provide a 5000L Milla Tank instead. This was an item participants did not have the option to directly

[3]Pictures of implemented projects can be found in appendix E.

Figure 2: Number of Messages by DTH



vote for in the DTHs, but one that reflected the participants' preference for improved water provision in affected wards.

3.3 Service Delivery Notification Calls

While our endline survey was conducted after all selected services were successfully delivered, we found it plausible that not all participants would be aware of the completed service project. Further, we worried that our inability to observe respondents' knowledge of project implementation would complicate the interpretation of our findings. For example, to what extent should a null (or perverse) effect be attributed to respondents' (mistaken) belief that services selected in the DTH had not been implemented? To address this, we made notification calls on behalf of the FCC to all treated units, informing participants the project chosen the the DTH had been successfully implemented. Note that by making these notification call to treated units but not control, we build the notifications calls into into our treatment.[4]

We successfully reached approximately 70% of treated units to inform them of the implemented services. These calls started in mid-November and were staggered across wards so that they started once service delivery was completed in that ward. The endline survey similarly was staggered and commenced after notification calls were completed, but never earlier than one week after delivery completion. While not part of the main intervention, we also made notification calls to a randomized subset of non-study property owners (described in section 9.2).

[4]Therefore, the treatment effect includes the heightened awareness of treated units regarding the implementation of the service.

4 Theoretical Model and Hypotheses

4.1 Primary and Secondary Hypotheses

Our primary hypothesis predicts that the Digital Town Hall intervention increases participant tax compliance.

H1: Invitation to DTH increases property tax compliance in 2022.

As secondary hypotheses, we also predict that treatment increases conditional and unconditional tax morale.

H2: Invitation to DTH increases unconditional tax morale.
H3: Invitation to DTH increases the belief that non-compliance can be justified if adequate services are not provided (service-conditional tax morale).

Our tax morale predictions may seem contradictory and therefore warrant some discussion. Consider two types of justifications that can be given for not paying taxes: (1) not paying taxes can be justified when services are not adequate; (2) not paying taxes can be justified on other grounds. We predict that treated respondents are more likely to agree with this service-based justification for not paying taxes, but less likely to agree with the non-service based justification for non-compliance.[5] Taken together, our predictions regarding conditional and unconditional tax compliance form an implicit prediction that treatment decreases respondents' acceptance of non-services justifications more than it increases their acceptance of service-based justifications for non-compliance.

4.2 Causal Mechanisms

In addition, we make four hypotheses about the mechanisms through which treatment impacts tax compliance and tax morale. We hypothesize that treated individuals will be more tax compliant and have higher levels of conditional and unconditional tax morale because we expect them to update positively about: (i) the amount of valued services they receive in return for tax payment ("fiscal exchange"), (ii) the opportunities to participate in politics and effect political change ("political efficacy"), (iii) the fairness and equity of the tax system ("fairness and equity"), and (iv) the level of enforcement ("enforcement"). This theoretical framework builds on the one presented in Prichard et al. (2019).

H4: Treatment increases willingness to expand fiscal exchange.
H5: Treatment increases political efficacy.
H6: Treatment increases perceptions of fairness and equity in the tax system.
H7: Treatment increases perception that non-compliance is punished.

4.3 Heterogeneous Effects

While we are primarily interested in the average treatment effect across respondents, we will also explore variation in the effect of the treatment on tax compliance and our measure of conditional tax morale. Below, we list the set of variables for which we investigate heterogeneous effects. Next to each variable, we specify our prediction about the direction of the treatment effect at higher levels of the variable.

For both tax compliance and conditional tax morale, we explore treatment effect heterogeneity with respect to:

[5] Non-service justification for non-compliance include: (i) the equity of the tax burden, (ii) one's precarious economic situation, (iii) being very busy, etc.

- Approval of Mayor (midline +; endline -)
- Co-partisan of Mayor (+)
- Perceptions of FCC institutional quality (-)
- Education (+)
- Political interest (+)
- Age (-)
- Tax liability (+)
- Fairness of tax assessment (-)
- Perceived neighbors' compliance (-)

Finally, in appendix B we outline a set of purely descriptive hypotheses on who actively participates in the DTH and who updates over service delivery.

5 Measurement

For our measure of tax compliance, we rely on FCC administration data, which allows us to observe individual level tax compliance behavior for the universe of property owners in Freetown. Our measure of tax compliance is a dummy variable equal to 1 if a property owner makes any tax payment in 2022. In 2021, the compliance rate in the control group was 25%.[6]

For our measures of conditional and unconditional tax morale, as well as our intermediate mechanisms outcomes, we rely on three rounds of survey data collected (i) prior to treatment assignment, (ii) post-treatment, but before the implementation of the selected service, and (iii) after implementation of the selected service.[7] Table 2 displays the number of completed surveys in each survey round. Survey measure summary statistics for the control group at baseline, midline, and endline can be found in Appendix A. Table 3 provides additional information about indicators (i.e., survey measures) and maps indicator onto the mechanism families outlined in section 4.

Survey round	Completed surveys	%
Baseline	3618	100 %
Midline	3304	91.3 %
Endline	2872	79.4 %

Table 2: Number of surveys per round

[6]Note that 16% of those who paid some tax in 2021 did not pay the full amount. For robustness, we can estimate the effect of the DTH on the percent of tax liability paid. For ease of interpretation, we prefer to use a binary compliance measure as the primary outcome.

[7]We provide financial incentives—packages of mobile data—for midline and endline survey takers to minimize attrition.

Indicator Name	Variables Description
Compliance Outcomes	
Tax compliance	A dummy variable equal to 1 if the property owner paid any tax in 2022, by the tax payment deadline of September 30. This information is taken from FCC administrative records.
Unconditional tax morale	A survey measure asking respondents whether "some people not paying the taxes that they owe" can always be justified, never be justified or something in between. Measured on a five-point Likert scale.
Conditional tax morale	A survey measure asking respondents whether they agree or disagree with the statement, "taxpayers could refuse to pay taxes if they are not receiving public services of adequate quality". Measured on a five-point Likert scale.
Mechanism Family 1: Fiscal Exchange	
Fiscal exchange willingness	A survey question that asks respondents if they agree with the statement that they would be willing to pay additional taxes to receive improved services. Measured on a five-point Likert scale.
Service satisfaction	How satisfied is the respondent with the Freetown City Council's provision of services? Measured on a five-point Likert scale.
Mechanism Family 2: Political Efficacy	
Opportunities for voice	Perceived opportunities for "citizens like you" to voice opinions about Freetown City Council operations. Measured on a four-point Likert scale.
FCC responsiveness	Perceived responsiveness of Freetown City Council to citizens' demands. Measured on a five-point Likert scale.
Participation efficacy	Perceived ability to participate in a political group. Measured on a five-point Likert scale.
Mechanism Family 3: Fairness of Taxation	
Tax system fairness	Agreement with the Freetown City Council's claim that the new property tax system is more fair. Measured on a three-point Likert scale.
Others' compliance	Estimated number of 10 closest neighbors receiving a tax bill that will pay.
Mechanism 4: Enforcement	
Punishment likelihood	Perceived likelihood that a non-compliant property owner will face legal consequences, assuming the Freetown City Council knows this person has not paid. Measured on a five-point Likert scale.

Table 3: Description of outcome / mechanism variables

6 Study Population and Sampling

To construct our sample frame we draw on FCC administrative records of the universe of taxable properties in Freetown, which contains a set of property characteristics and property owner contact information. To be eligible to participate in the Digital Town Hall a property owner must (i) own a property in one of the 30 study wards and (ii) have WhatsApp on their phone. For property owners that own multiple properties, we coded them as being eligible for the DTH in the ward that contains their highest value property (i.e., highest tax fee).[8] We used owner contact information in FCC's administrative records to call 15,977 property owners in our study wards. We refer to this set of property owners as the "call list". From the call list we were able to confirm 4,860 property owners that had WhatsApp on one of their phones; these property owner were eligible to be selected into the Digital Town Hall intervention.

The set of 15,977 property owners on the call list is *not* a random sample of property owners from the 30 study wards. First, (most) property owners on the call list own properties of above median value. As a response to COVID-19, the FCC intended to waive property tax for 2020 on properties of below median value. As our intervention was originally scheduled for early 2020, it was necessary to target the DTH intervention at property owners who owned properties above the median property value. Politics related to the tax reform caused us to delay the DTH intervention until early 2021. However, during the calling process we unintentionally verified 450 property owners who own a property below the median value. We included these property owners in our sample. Second, in a previous version of our research design, we planned to allocate treatment status using a two-stage randomization procedure, to mitigate and estimate geographic spillover (as in Sinclair et al. 2012). Under that research design, properties were divided into geographic clusters using a grid overlay and properties within five meters of the edge of a grid cell were ineligible for the study. We constructed the call list with this research design in mind, thereby removing properties within five meters of the grid cell edge.

We were able to complete baseline surveys with 3859 of the 4860 verified property owners (79.4%). To mitigate spillover, we drew a restricted sample from this set of property owners such that each property is at least 15 meters from the closest study property. The restricted sampling leaves us with a final sample of 3619. Figure 7 in appendix C displays the distribution of the distance from each property to the closest property in the sample.

In the next iteration of this document, we will use data from FCC administrative records to compare property owners in our study sample to non-study property owners in Freetown.

7 Treatment Assignment

We assign treatment status using a matched-pair design, leveraging baseline data to match similar observations into groups of two. We create 1809 pairs and then assign one observation in each matched-pair to treatment and the other to control.

7.1 Matching

We match property owners using the following covariates:

[8] This prevents the same property owner from being assigned to the DTH in multiple wards or being assigned to both treatment and control conditions. We made multiple property owners eligible for the DTH in the study wards where their highest value property was located as we reasoned that they were more likely to be resident of these properties and more likely to be involved in the administering of these properties (and therefore more likely to be directly involved in the decision to pay property tax). Note that there are only a handful of DTH participants who own multiple properties and are in the DTH of their second highest value property. In these instances, the ward in which they have a higher value property is not a study ward.

- Unconditional tax morale
- Service conditional tax morale
- Perceived probability of punishment for non-compliance
- Satisfaction with FCC service provision
- Tax reform awareness and support
- RDN received in 2019 or 2020
- Opportunities to voice opinion about FCC governance
- Willingness to believe member of opposing party
- Mayor approval
- FCC councilor approval
- Gender
- FCC responsiveness
- Age
- Property value
- Education

We generate matched-pairs using the *blockTools* package in R. We use the Optimal Greedy ("opt-Greedy") matching algorithm to find best matches along mahalanobis distance. In this matching process we weight certain variables higher than others, in line with our expectations that certain variables are a stronger predictor of our outcomes of interest. We place the greatest weight on our measure of unconditional tax morale—we expect this to be the strongest predictor to tax compliance, in line with the common use of this variable as proxy for tax compliance behavior. We place equal weight on another set of six measures from our baseline survey. Three of these measures are important factors in the literature on tax compliance: (i) service conditional tax morale, (ii) perceived likelihood of punishment for non-compliance, and (iii) satisfaction with FCC service provision. We also place equal weight on the (iv) gender of the property owner, (v) their awareness and support of the property tax reform,[9] and (vi) the number of these five variables that were imputed.[10]

Variable name	**Weights**	**Mean**	**SD**	**Min**	**Max**	**n imputed**
Unconditional tax morale	1.10	3.77	1.55	1.00	5.00	25
Service conditional tax morale	1.00	1.96	0.96	1.00	3.00	11
Perceived probability of punishment	1.00	4.06	1.11	1.00	5.00	52
Satisfaction with FCC service provision	1.00	3.64	1.17	1.00	5.00	35
Gender (female = 1)	1.00	0.31	0.46	0.00	1.00	0
Reform awareness / support	1.00	2.38	0.67	1.00	3.00	19
RDN delivered 2019 or 2020	0.90	0.83	0.38	0.00	1.00	0
Opportunities for voice	0.10	2.13	0.99	1.00	4.00	174
Mayor approval	0.10	4.23	0.89	1.00	5.00	79
Councilor approval	0.10	2.73	1.22	1.00	5.00	122
FCC responsiveness	0.10	3.17	1.19	1.00	5.00	199
Believe opposition member	0.10	3.00	1.55	0.00	5.00	132
Age	0.09	51.77	12.93	20.00	100.00	11
Property tax value (USD)	0.09	60.25	87.45	2.88	1281.85	0
Education [0-2]	0.09	1.31	0.62	0.00	2.00	259

Table 4: Summary statistics of matching variables

[9] We create a three level ordinal variable based on two survey items. A first group consists of respondents who have heard of the reform and strongly/somewhat support it; a second group consists of respondents who (a) have heard of the reform and feel neutral towards it and (b) have not heard of the reform; a third group consists of respondents who have heard of the reform and somewhat/strongly oppose it.

[10] This avoids matching observations with missing values on these key variables to observations that have non-missing values close to the mean.

Table 4 presents descriptive statistics and match weights for our matching variables. If a respondent refused to answer a question or said they "did not know" we imputed the value as the unconditional mean of the variable.[11] The last column displays the number of observations that were imputed for matching. Note that in general, the number of imputed responses is low.

7.2 Treatment Assignment and Balance

In each matched pair, one unit is assigned to treatment and one to control. We implement this randomization in R using the *block_ra* function in the *randomizr* package. Table 5 presents balance statistics for outcome variables (at baseline), matching variables, and several property-level characteristics.[12] Across this range of variables, differences in means between treatment and control group are small. Figure 3 visualizes treatment assignment across Freetown.

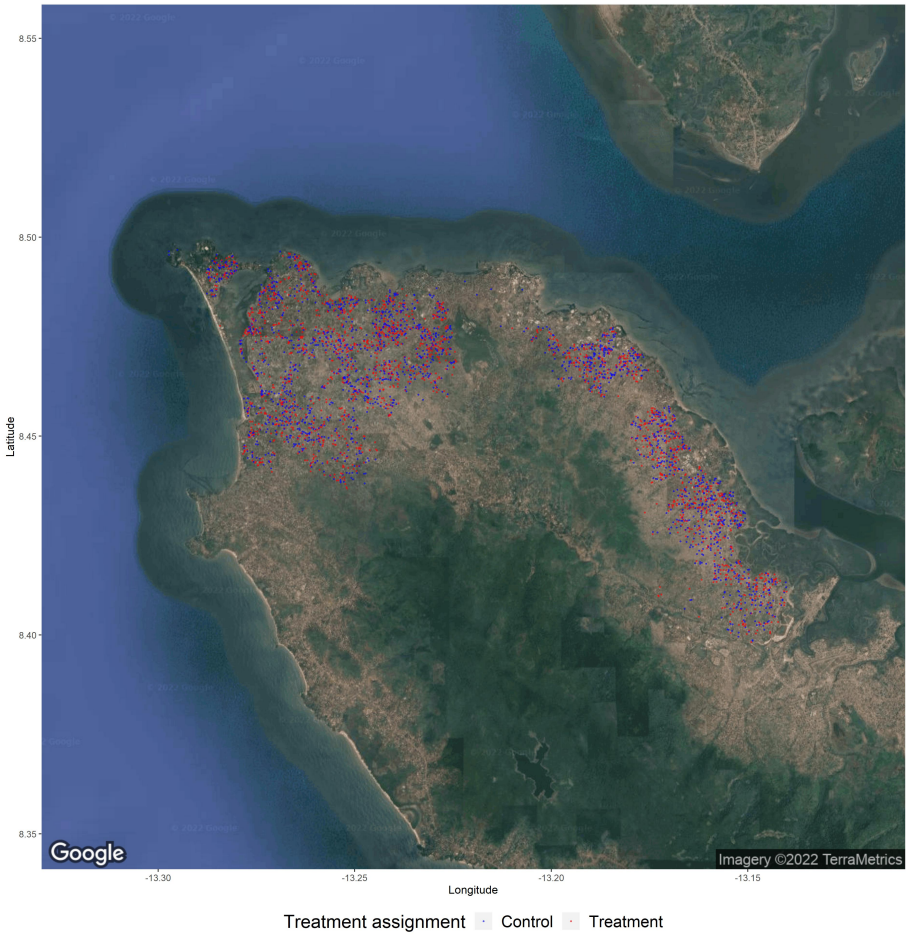


Figure 3: Digital Town Hall treatment assignment in Freetown (red = treatment)

[11] Following suggestions of: https://egap.org/resource/10-things-to-know-about-missing-data/

[12] The last column of Table 5 standardizes the difference relative to the standard deviation of the control group.

Variable name	Mean (Z0)	Mean (Z1)	SD (Z0)	N (Z0)	N (Z1)	Raw dif	Std. dif
Fiscal exchange willingness	4.19	4.18	1.22	1805	1804	0.01	0.008
Political efficacy	1.76	1.74	1.14	1794	1793	0.02	0.018
Tax system fairness	2.12	2.11	0.79	1112	1129	0.01	0.013
Neighbors' compliance	5.13	5.07	2.41	1138	1105	0.06	0.025
Unconditional tax morale	3.77	3.78	1.55	1797	1799	-0.01	-0.006
Service conditional tax morale	1.96	1.96	0.96	1805	1803	0.00	0.000
Perceived probability of punishment	4.06	4.06	1.11	1788	1781	0.00	0.000
Satisfaction with FCC service provision	3.64	3.64	1.17	1790	1796	0.00	0.000
Opportunities for voice	2.12	2.13	1.00	1719	1736	-0.01	-0.010
Mayor approval	4.23	4.22	0.89	1770	1774	0.01	0.011
Councilor approval	2.73	2.74	1.22	1751	1751	-0.01	-0.008
FCC responsiveness	3.17	3.17	1.18	1712	1719	0.00	0.000
Gender (female = 1)	0.31	0.30	0.46	1809	1809	0.01	0.022
Age	51.65	51.88	13.00	1803	1804	-0.23	-0.018
Reform awareness / support	2.38	2.37	0.67	1794	1806	0.01	0.015
Property tax value (USD)	60.12	60.38	86.49	1809	1809	-0.26	-0.003
Believe opposition member	3.02	2.99	1.55	1749	1744	0.03	0.019
Received RDN 2019 or 2020	0.83	0.83	0.38	1809	1809	0.00	0.000
Education [0-2]	1.30	1.32	0.62	1685	1694	-0.02	-0.032
Tax compliance 2020	0.07	0.07	0.25	1809	1806	0.00	0.000
Number of properties with tax liability 2021	1.93	1.89	1.48	1809	1806	0.04	0.027
Property on pave road	0.27	0.25	0.44	1809	1806	0.02	0.045
Property has water	0.47	0.47	0.50	1809	1806	0.00	0.000
Property has drainage	0.36	0.36	0.48	1809	1806	0.00	0.000

Table 5: Balance table

8 Data Analysis

In this section we describe our approach for estimating the effects of the Digital Town Hall on two types of outcomes: (i) tax compliance behavior and (ii) survey-based measures. The nature of our intervention allows for one-sided noncompliance and indeed not all property owners who were invited to join the DTH actually joined it. Of the 1809 property owners assigned to treatment, 1459 (80.7%) joined WhatsApp groups of the DTH.[13] Therefore, we will provide Intent-to-Treat (ITT) and Complier Average Casual Effect (CACE) estimates. While ITT estimators provide unbiased estimates of being assigned to treatment, the presence of one-sided non-compliance means that our ITT estimate will underestimate the effect of *participating* in the DTH. We note, however, that our CATE estimates may be upwards biased due to a potential exclusion restriction violation, which stems from notification calls going to all treated units, not just treated units that joined the DTH. Given the merits and weaknesses associated with each estimator, we consider estimates from both estimators when discussing results.

In sections 8.1, we discuss our estimation of ITTs. Section 8.2 discusses our estimation CACEs and outlines the assumptions required for making unbiased estimates of this quantity. In section 8.3, we discuss inference and section 8.4 discusses estimation of heterogeneous effects. In section 11, we provide results from simulation exercises that estimate minimal detectable effects for each outcome of interest.

8.1 Intent-to-Treat Effect (ITT)

For both compliance and survey-based measures, we estimate ITTs using the following equation:

[13] In section 3.1.1 we noted that 1,637 of 1,809 respondents who were invited to join the DTH consented to join it. However, 159 respondents who consented to join the DTH never joined the WhatsApp chat group.

$$Y_{ijt_2} = \beta_1 DTH_i + \gamma Y_{ijt_1} + \sum_{j=1}^{1809} \theta_j PAIR_j + \delta_w + \lambda \mathbf{X}_i + \epsilon_i \quad (1)$$

Where Y_{ijt_2} is the endline (t_2) outcome of individual i in pair j; DTH_i is an indicator variable equal to 1 if owner i is assigned to treatment and β_1 captures the average treatment effect of the Digital Town Hall; Y_{ijt_1} is the baseline outcome for owner i in pair j. When Y is property tax compliance behavior, Y_{t_1} refers to tax compliance behavior in 2020; When Y is a survey outcome, Y_{t_1} refers to the baseline survey outcome. $PAIR_j$ is an indicator variable equal to 1 if owner i belongs to pair j; δ is a vector of ward fixed effects and ϵ_i is the error term.[14]

$\mathbf{X}$ is a set of property-level characteristics that we include for covariate adjustment only when estimating treatment effects on property tax compliance behavior.[15] We include the following control variables: (i) log total tax liability, (ii) number of properties with any liability, (iii) access to water, (iv) access to drainage, (v) property in an informal settlement, (vi) property has fencing or gate, (vii) property has garage, (viii) street condition, (ix) street type (x) ease of property access, (xi) window quality, (xii) tax bill type received. Where covariate data is missing, including baseline values of the outcome of interest, we impute missing data using the baseline mean of that variable. Appendix D displays summary statistics for control variables.[16] While equation 1 is our primary specification, we also analyze the data without ward fixed effects and without the control variables in $\mathbf{X}$ as robustness checks.

Note that the matched pair fixed effects term in equation 1 implies that where outcome data is missing, both observations in the matched pair are effectively dropped from the estimation, maintaining unbiased estimates (King et al. 2007; Fukumoto 2015).[17] While we do not have missingness on our primary outcome (tax compliance behavior), we will have missing outcomes for survey-based measures, either due to survey item non-response or respondent attrition in our midline and endline surveys.[18] Dropping both observations in a pair when one observation is missing maintains unbiased estimates at the cost of statistical power by throwing out non-missing values. As a robustness check, we also analyze the data without dropping non-missing observations. To do this, we re-match non-missing observations who are in a pair where the paired observation is missing, using the same matching strategy as used in the original randomization (see section 7.1). These estimates are unbiased under the assumption that patterns of missingness are independent of treatment assignment. That is, the units that attrit must have similar potential outcomes in treatment and control and the true value of missing survey items must be equivalent in treatment and control. While this is an assumption that cannot be directly tested, we will analyze patterns of missingness to shed light on the plausibility of this assumption. First, for each survey item, we compare the rate of missingness across treatment and control. If rates of missingness are similar across groups, the above assumptions are more plausible. Second, because there is not attrition in our baseline survey, we can check for balance of prognostic covariates among the respondents who attrit in midline and endline survey. If we find balance on prognostic variables the above assumptions are more plausible.

[14] We prefer to condition on the baseline value—rather than use the change score as the dependent variable—because the randomized design limits the risk of confounding bias. See this excellent Declare Design blog post.

[15] Equation 1 controls for survey-based outcomes that we expect to predict compliance and survey outcomes through the inclusion of matched-pair dummies (see section 7.1 for details).

[16] For each covariate, we include a dummy variable that indicates if the covariate value is imputed.

[17] Note that if there is treatment effect heterogeneity and missingness is non-random, estimated ATEs may not equal the true sample ATE. With this strategy the estimand changes to the average treatment effect among the subset of the sample who are in a pair where both values are non-missing. The estimate of this quantity is unbiased.

[18] Recall that we observe tax compliance through administrative data and therefore do not anticipate missingness.

8.2 Complier Average Causal Effect (CACE)

To estimate the effect of a property owner joining the town hall—rather than being assigned to treatment—we use an instrumental variable regression framework. In this set-up, joining the DTH is conceptualized as the treatment and our invitation to property owners to join the DTH is conceptualized as the instrument (or encouragement). This estimator captures the local ATE among the set of people who comply with treatment, which in this case refers to property owners who are invited to join the DTH and consent to join the DTH.

This estimate is causally identified when the following assumptions hold. First, there should be no direct effect of the instrument on outcomes of interest, which in the context of our study refers to an effect of the invitation to join the DTH. Second, there can be no indirect effect of the encouragement, which in our context would mean that being invited to the DTH affects the outcome of interest through some intermediate factor. Third, the "no defier" assumption posits that an invitation to join the DTH should not lower the probability that a property owner consent to join the DTH. This possibility is ruled out by our study design, as property owners can only join the DTH if they have been invited. Fourth, the instrument should not be correlated with potential outcomes. In our design, the instrument is randomly assigned, ruling out this possibility. Fifth, the instrument should be "relevant," in the sense that is has a substantively meaningful impact on the probability to take up treatment. This assumption is justified in our study, as the invitation to join the DTH increases treatment uptake from zero to 80.7%.

The most plausible threat to the unbiasedness of our CACE estimates is that notification calls were made to all participants who were assigned to treatment, not just those who joined the DTH; these calls were not made to control units (see section section 3.3). The exclusion restriction would be violated if these notification calls impact outcome of interest, as this would mean the instrument affects outcomes of interest through channels other than the treatment of participating in the DTH. How should we handle this possible violation to the exclusion restriction? A back of the envelope calculation reveals that the size of the bias in the CACE estimator will be about a quarter of size of the effect of the notification call.[19] While we cannot estimate the effect of the notification calls in our sample, we can estimate the effect of these notification calls on compliance behavior for a sample of property owners not involved in the study (see section 9.2). If these estimates are near zero, we have good reason to believe that our CACE estimates are unbiased; if these estimates are distinguishable from zero, we have an idea of the direction and magnitude of the bias in the CACE estimates. Note that we can only estimate the impact of the notification calls on compliance behavior, as we do not have survey outcomes for this sample.

Less concerning, but still plausible, is the possibility that the encouragement itself affects outcomes of interest. Property owners that were assigned to receive an invitation to the DTH were contacted by our research team on behalf of the International Growth Centre (IGC). A member of the research team briefly explained the features of the DTH and asked if the participant would give their consent to join. In asking the property owner for consent, the research assistant noted that the DTH groups are *"organized by the Mayor of Freetown and your Ward Councilor"* and that property owners are *"invited to interact with your Councilor and the Mayor, to share your views on development projects in your ward and in Freetown, and then have a direct vote on how some development funds that have been allocated to your ward will be spent"*. It is conceivable that the invitation to join the DTH has an affect on attitudes and future behavior. If this was the case, our CACE estimates of the effect of participating in the DTH would be upward biased.

[19] Recall the true effect of the DTH β_{DTH} and the effect of the notification call β_{bias}. Given the compliance rate of .807, the ITT estimate is $.807 * \beta_{DTH} + .193 * \beta_{bias}$. We get the CACE estimate by dividing this quantity by the compliance rate of .807, such that $CACE = \beta_{DTH} + .239 * \beta_{bias}$.

8.3 Inference

We will report estimates with heteroskedasticity-robust standard errors (HC2). As randomization occurs at the level of the observation (property owner), we do not cluster standard errors. We calculate p-values using randomization inference. Several of our mechanism families contain multiple indicators. Therefore, in addition to presenting standard RI p-values on each mechanism outcome, we also report p-values adjusted for multiple comparisons. Specifically, we make the Benjamini-Hochberg correction to constrain the false discover rate (FDR) at .05 using the two-step adjustment detailed in Anderson (2008). We make adjustments within mechanism families; we do not adjust p-values across mechanism families.

8.4 Heterogeneous Effects

To estimate heterogeneous effects, we will first recode (where necessary) variables of interest (see section 4.3) into binary variables. Levels for each variable will be coded as follows:

- Approval of Mayor: (1 = "strongly approve" or "approve"; 0 = "strongly disapprove" or "disapprove" or "in the middle").
- Political partisanship: (1 = APC supporter; 0 = non-APC supporter).
- Perceptions of FCC institutional quality:[20] (1 = above median index; 0 = below median index).
- Education: (1 = post high school education; 0 = high school or lower).
- Political interest (1 = above median; 0 = below median).
- Age: (1 = above median; 0 = below median).
- Tax liability: (1 = above sample median; 0 = below sample median).
- Fairness of tax assessment: (1 = "unfair"; 0 = "fair" or "somewhat fair").
- Perceived neighbors' compliance: (1 = $\geq$ 5 of 10 neighbors; 0 = $\leq$ 5 of 10 neighbors).

We then estimate heterogeneous treatment by adding each variable (in separate regressions) to the main estimating equation, interacting it with the treatment indicator.

8.5 Connection to Analyses in Other Projects

In a (currently) separate project, we investigate the effectiveness of DTHs as tools for strengthening political accountability during COVID-19.[21] In that project, our main outcome families are (i) attitudes towards political representatives and political institutions, (ii) political knowledge and efficacy, and (iii) community cohesion and interpersonal trust. The survey indicators used to construct indices for these outcome families overlap with some of the survey indicators we use in this project on tax compliance. For this other project on political accountability during COVID, we analyzed the effects of the DTH on midline survey outcomes; we have not yet analyzed any endline data collected for the DTH project. While we note the above for transparency reasons, we do not believe that the analysis of midline outcomes impacted our analysis plan for the tax compliance study (the subject of this PAP) in any meaningful way. Lastly, we leave open the possibility that these two sets of analyses will be combined in the future.

9 Unbundling the Intervention

As described in section 3, the intervention under study contains three components: (i) digital town halls, (ii) delivery of selected services, and (iii) notification that selected services were delivered. While only property owners in the treatment group participated in the DTH and were notified through a

[20] A z-score index created from four variables: (i) FCC responsiveness, (ii) FCC efficiency, (iii) FCC corruption, and (iv) FCC punishment likelihood, where missing values are imputed using the group mean.

[21] The pre-analysis plan for that project can be found at this link: https://osf.io/cg738.

phone call that selected services were implemented, property owners in both treatment and control had services implemented in their ward (see table 1). This implies that the estimand in our primary analysis is the effect of participating in a digital town hall, conditional on services being delivered and having knowledge that services were delivered. In additional analyses, we attempt to further unbundle the effect of different components of the intervention and estimate two additional quantities of theoretical and practical interest:

(i) the effect of the DTH conditional on services *not yet being delivered*.
(ii) the effect of being notified that services were implemented.

In section 9.1, we attempt to parse out how the effect of the DTH varies conditional on the implementation and knowledge of the services selected in the DTH. In section 9.2, we describe a strategy for estimating the effects of *notifying* property owners about implemented services. These estimates provide useful benchmarks for interpreting our main results about the effects of participating in DTHs.

9.1 How Does Service Delivery Condition the Effects of the DTH?

Participatory budgeting programs, like the Digital Town Hall, may not lead to changes in attitudes and behaviors if participants are uncertain about the government's willingness or ability to provide the selected service. Therefore, we might expect that the magnitude and plausibly direction of the effect of the DTH is conditional on the delivery of services selected within it. An ideal way to estimate these conditional effects would be to randomly assign some DTHs to have their selected service implemented—leaving a second set of town halls where the selected service was not implemented—and then compare outcomes across these two groups. In our study, all selected service were implemented, so this comparison is not possible.

As a second best strategy, we attempt to estimate these conditional effects by leveraging multiple rounds of survey data: we conducted one round of data collection after the Digital Town Hall but before service delivery (midline) and conducted a third round of survey data collection after service delivery (endline). One (naive) approach is to compare two point estimates from two different regressions: one where the dependent variable in equation 1 is the endline outcome and a second where the dependent variable in equation 1 is the midline outcome. In this approach, we interpret the point estimate from the midline survey as the effect of the DTH conditional uncertainty regarding service delivery and we interpret the point estimate from the endline survey as the effect of the DTH conditional on service delivery. The major weakness with this approach is that inferences about the difference between the effect of the DTH conditional on service delivery will be confounded by the possibility that treatment effects on attitudinal outcomes deteriorate over time.[22]

9.2 Notification Calls

Within our study sample, we called treatment property owners to notify them about the implementation of services selected through the DTH. As this information was provided only to treatment units, we cannot cleanly separate out the effect of this information alone from the rest of the other components of the intervention. To get the pure effect of information about services, we randomized information about the implementation of services to property owners outside of our study sample.

We constructed a sample frame of 15,217 non-study property owners that met the following criteria: (i) own a property in one of the 30 study wards, (ii) have a phone number on file at FCC, (iii) had not been called as part of the initial verification process that selected property owners into the study,[23](iv)

[22]For example, even if the effect of the DTH is larger conditional on service delivery, we may observe no difference between point estimates made with midline and endline data because treatment effects on attitudes may erode with time from the DTH.

[23]That is, not part of the call list described in section 6.

and had not paid taxes in either 2020 or as of October 23, 2021.[24]

Those 15,217 property owners were randomized into treatment and control groups—where treatment is defined by receiving a call notifying the property owner of implemented services—blocking on tax rate decile within each ward. Treatment non-compliance in this context is failure to speak to a treated property owner after repeated calls. Given our expectation of high rates of non-compliance, we randomized 8,951 owners into treatment 6,266 to control.

Treated respondents receive a call from a surveyor, who identifies themselves as calling on behalf of the FCC. After confirming the respondent's personal information, the surveyor provides the respondent with the following information: *"Recently, in your ward [WARD NAME], [PROJECT DESCRIPTION] has been built by a construction firm on behalf of the Freetown City Council. This is at [PROJECT LOCATION]. This project was funded by resources associated with the FCC's property tax reform."* Surveyors then ask the property owner if they have heard of this project, and if so, if they had visited it. Then, surveyors conclude the call with the following text: *"We're looking forward to continuing to work with people in your community to better understand the most pressing local development needs. This is one of the steps the FCC is taking to develop the city as part of the FCC's ambitious plan to Transform Freetown. If you have any further questions about the project in your ward, you may contact us at the following phone numbers: XXX or XXX."*

We will estimate the effect of receiving a service notification call from the survey team using an instrument instrumental framework, where the instrument is the randomly assigned treatment and the endogenous treatment is a indicator variable equal to 1 if the property owner answered the notification phone call. As these are non-study properties for which we do not have survey data, our outcome of interest is tax compliance.

10 Spillovers

In this project, spillovers are a potential threat to inference. Spillovers may bias our estimates of treatment effects in two ways. First, if treatment spills over to control units we underestimate treatment effects because we no longer observe control potential outcomes for (some) control units. Second, if treatment spills over to non-study units (i.e., property owners not in our study) failing to estimate spillovers means some indirect treatment effects are not accounted for. At a theoretical level, the direction of the bias introduced by spillovers is ambiguous. On the one hand, those uninvited may feel disappointment or frustrated with their inability to participate, lowering control potential outcomes and upward biasing our estimate. On the other hand, even uninvited property owners might appreciate that political representatives are making costly investments into soliciting citizen input in challenging times. Moreover, the implemented services are local public goods intended to profit the broader community and even uninvited citizens should update positively after receiving more services. Therefore, we believe that positive spillovers are more plausible, which would downward biases our estimates.

Spillovers are also of substantive interest. Politicians making costly investments into engaging citizens through participatory budgeting Town Halls seek to receive credit and tax dollars, even from citizens who were not invited to the Town Hall. Yet, as López-Moctezuma et al. (2022, p.20) in a recent high-profile publication note, disentangling "the effects of town-hall meetings on attendees from the spillover effects to untreated voters" has not yet been convincingly done and remains a "future avenue of fruitful research." While evidence on the spillover effects of participatory budgeting (our independent variable) is thin, there is a substantial literature demonstrating that tax compliance behavior (our primary dependent variable) can spillover under certain conditions. Boning et al. (2018) demonstrate that IRS audits influence the compliance behavior of firms that are geographically proximate or share

[24] Recall that 2020 tax compliance was only about 3%.

the same tax preparer of the one visited. In contrast, Meiselman (2018) randomizes the content of tax messages in the city of Detroit and does not find evidence that units living closer to a taxpayer treated with the enforcement message are more likely to comply. Drago et al. (2015) also employ a messaging experiment (here in small municipalities in Austria) and finds significant geographic spillovers.[25] In contrast to our study, this literature focuses on enforcement interventions.

Spillovers might be a function of either geographic or social proximity. Our experimental design includes some safeguard against geographic spillovers—we included a geographic buffer of at least 15m between study units, allowing us to exclude neighboring properties and apartments in the same building. In section 10.1, we lay out a design-based strategy to detect geographic spillovers. In section 10.2, we describe our procedure for detecting social network-based spillovers.

10.1 Estimating Geographic Spillover

We use a design-based strategy to estimate spillovers that occur due to geographic proximity between properties. For this analysis, we focus on tax compliance spillovers from treated properties to properties outside of our study.[26] Using administrative data, we observe compliance outcomes for 95,769 properties that are not eligible for the intervention, which we refer to as "non-study properties".[27] Our approach compares non-study properties geographically proximate to treated study properties to non-study properties proximate to control study properties.[28] Under this approach, the probability that a non-study property is indirectly treated is correlated with the building density in the area where that non-study property is located. This means that point estimates from unweighted regressions and conventional standard error-based p-values may be biased. As discussed further below, we use inverse propensity score weighting (IPW) and randomization inference (RI) to address these issues.

We estimate spillovers with the following equation:

$$Y_{i_{2022}} = \beta_1 SPILL_i + \gamma Y_{i_{2020}} + \lambda \mathbf{X}_i + \delta_w + \epsilon_i \tag{2}$$

Where $Y_{i_{2022}}$ is the binary tax compliance outcome of non-study property owner i in 2022; $SPILL_i$ is a dummy variable equal to 1 if there is at least one treated study property 64 meters or less from non-study property owner i. Therefore, β_1 captures the spillover effect on tax compliance of being close to a treated property owner. $Y_{i_{2020}}$ is the tax compliance behavior of property owner i in 2020; δ is a vector of ward fixed effects; $\mathbf{X}$ is the set of property-level characteristics described in section 8.1, included as covariate adjustment. We also estimate spillovers where tax compliance in 2021 is the dependent variable.

[25] Using the criterion of a geographic household distance of 50m or less to identify network members yields a higher tax compliance likelihood of untreated network neighbors of treated units of between 5 and 7 percentage points. They no longer detect spillovers when they increase the distance to 500m—suggesting that spillovers are geographically confined. Spillovers in their study are strongest when the intervention is received by an individual with high eigenvector centrality.

[26] We do not pre-register strategies to estimate (i) spillovers from treatment to control properties in our study or (ii) spillover effects for attitudes or behaviors other than tax compliance.

[27] Recall that some individuals own multiple properties. Intuitively, the effects of the DTH should only spillover to affect the compliance behavior of a proximate non-study property when the property owner is living there. As we lack data on the residence of property owners who own more than one property, we assume that these multiple property owners are living in their highest value property. Therefore, our spillover analysis is restricted to the set of 74,352 non-study properties that are the highest value property registered to a given property owner.

[28] See Miguel and Kremer (2004) for an example of a (prominent) study that uses non-experimental units (i.e., units that are not themselves part of the randomization) to estimate spillovers.

As the density of buildings varies across the city, the probability of being assigned to "spillover treatment" (i.e. the probability that $SPILL_i$ is equal to one in equation 2) varies across properties. That is, non-study properties in denser areas are more likely to be assigned to spillover treatment because they are more likely to be close to more study units. In this context, unweighted regressions can be biased because building density (and therefore treatment assignment) may also be correlated with compliance behavior.[29] To address this, we weight observations by the inverse probability of being assigned their spillover treatment condition, where assignment probability is calculated by re-simulating treatment assignment of study properties (Blattman et al. 2021; Gerber and Green 2012; Chen et al. 2010). Note that this implies non-study properties more than 64 meters to the closest study property are weighted zero (i.e., not used to calculate spillover effects).

When treatment assignment probability does not map perfectly to observable characteristics, the potential correlation between treatment assignment probabilities and potential outcomes also creates inferential challenges. As noted by Blattman et al. (2021), "Thus there is no geographic unit with which to calculate clustered standard errors." Instead, we follow these authors and calculate randomization inference p-values, based on 10,000 simulated randomizations under the assumption of no effect.

Clearly, estimating equation 2 requires that we define a distance D at which a non-study property is "close" to a study property and therefore subject to spillovers. How did we arrive at 64 meters as our preferred value of D? As our goal is to uncover spillover effects if they exist, (rather than estimating the average spillover effect in some pre-defined ring around treated units) selecting a value of D involves a bias-variance trade off. On the one hand, selecting a higher value of D increases the number of properties used to estimate the parameter of interest and lowers variance in our estimate. On the other hand, if spillover effects decay with distance, selecting higher values of D lowers the average treatment effect, making effects harder to detect. Absent a theory-driven procedure for selecting D, we opt for a pragmatic approach. While the overall number of non-study properties used in the spillover estimation increases with higher values of D, the number of spillover control units is maximized when D equals 64 meters. Values of D greater than 64 have increasing units in the spillover treatment condition, but decreasing units of spillover control units. Given that the motivation for selecting higher values of D is to increase precision, selecting a value of D greater than 64 meters requires that the loss of precision brought on by the decline of units in the control arm is outweighed by increase in precision due to additional units entering into the treatment arm. When D is equal to 64 meters the treatment spillover arm has 24,177 units, compared to 10,637 units in spillover control; therefore, we privilege maintaining control units over gaining treatment units.[30] Substantively, setting D equal to 64 meters implies a spillover zone of roughly four or five properties in each direction from the treated property, in accordance with the idea that treatment spills over between neighbors. As a robustness check, we will also present a coefficient plot where we estimate spillover effects at a range of values for D—20 to 100 meters, in 10 meter increments—plotting estimates against the corresponding D value.

10.2 Detecting Social Spillovers

But what about spillovers that do not occur due to geographic proximity? As our intervention relies on WhatsApp, it could be that information was shared through WhatsApp or other messaging services along non-geographic network lines. We utilize our midline and endline surveys to get at these spillovers.

Our main approach is to provide descriptive evidence on the frequency with which spillovers between treatment and control units occurred, leveraging items from the midline and endline survey. For our primary strategy for detecting social spillover, we ask control respondents if they know someone

[29] Imagine, for example, potential differences in compliance behavior between densely packed informal settlements and spacious affluent neighborhoods.

[30] Figure 11 in Appendix F displays the number of non-study units in each spillover condition used in estimation at different values of D.

who participated in the DTH and then if they say yes, we ask the identity of that person. We then cross-reference this claim. This provides us with an estimate of the number of control units who at least know someone participating in the DTH. We also directly ask control respondents if they are aware of the DTH. If so, we ask respondents what exactly they know about the topic (open-entry) and how they learned about it. We then ask for their immediate reaction when they first heard of the topic to get at the direction of bias spillovers would introduce. In addition, our survey contains a module that includes other questions that get at spillover potential including: (i) the frequency of WhatsApp usage, (ii) the number of WhatsApp contacts, (iii) the number of WhatsApp contacts the respondent communicated with in the last month. We also directly ask treated respondents how many non-household members and non-DTH participants in Freetown they talked about the DTH with.[31]

In terms of estimation, we cannot rely on design-based inference as we lack an exogenous measure of network size (e.g., the number of contacts of owner i that were eligible for receiving treatment). To complicate things further, our survey measures of WhatsApp network size are measured post-treatment. Therefore, estimating the compliance outcome of control unit i as a function of network size indicators and a dummy of whether they have correctly named a treatment individual, rests on the assumption that treatment status did not affect network size. Compared to the geographic spillover estimation above, we are less likely to detect spillovers here also because the following estimation is significantly less powered as only the approximately 1,800 control status owners are being included. Specifically, we estimate:

$$Y_i = \alpha + \beta_1 SPILL_{social\,i} + \beta_2 FrequencyUsage_i + \beta_3 NetworkSize_i + \beta_4 FrequencyContact_i + \epsilon_i \quad (3)$$

Where Y_i is the binary tax compliance outcome of property owner i in the control condition; $SPILL_{social}$ is a dummy variable equal to 1 if at least 1 treated study property owner was correctly identified; β_1 captures the effect on tax compliance behavior of correctly identifying a DTH participant; ϵ_i is the error term. $FrequencyUsage$ is an ordinal variable on how often a participant has used WhatsApp over the last seven days. $NetworkSize$ is the total number of self-identified contacts a participant has in WhatsApp. $FrequencyContact$ is the total number of contacts a participant has in WhatsApp that the participant had contact with over the last month.

11 Power Simulations

In this section, we provide power simulations for two analyses: ITT and heterogeneous effects. Simulations use mock data, where we generate control and treatment potential outcomes for each unit. For control units, control potential outcomes are true observed values; for treated units, control potential outcomes are generated from control group data. A treatment potential outcome is then generated for each unit by applying an increase of one scale point to control potential outcomes, with varying probabilities. Treatment potential outcomes are bounded by the limits of our measurement scale; that is, if an observation has a control potential outcome of 5 on a 5 point scale, the treatment potential outcome will also be 5. We analyze this data with the specifications outlined in section 8.

11.1 Power Simulations: ITT Analysis

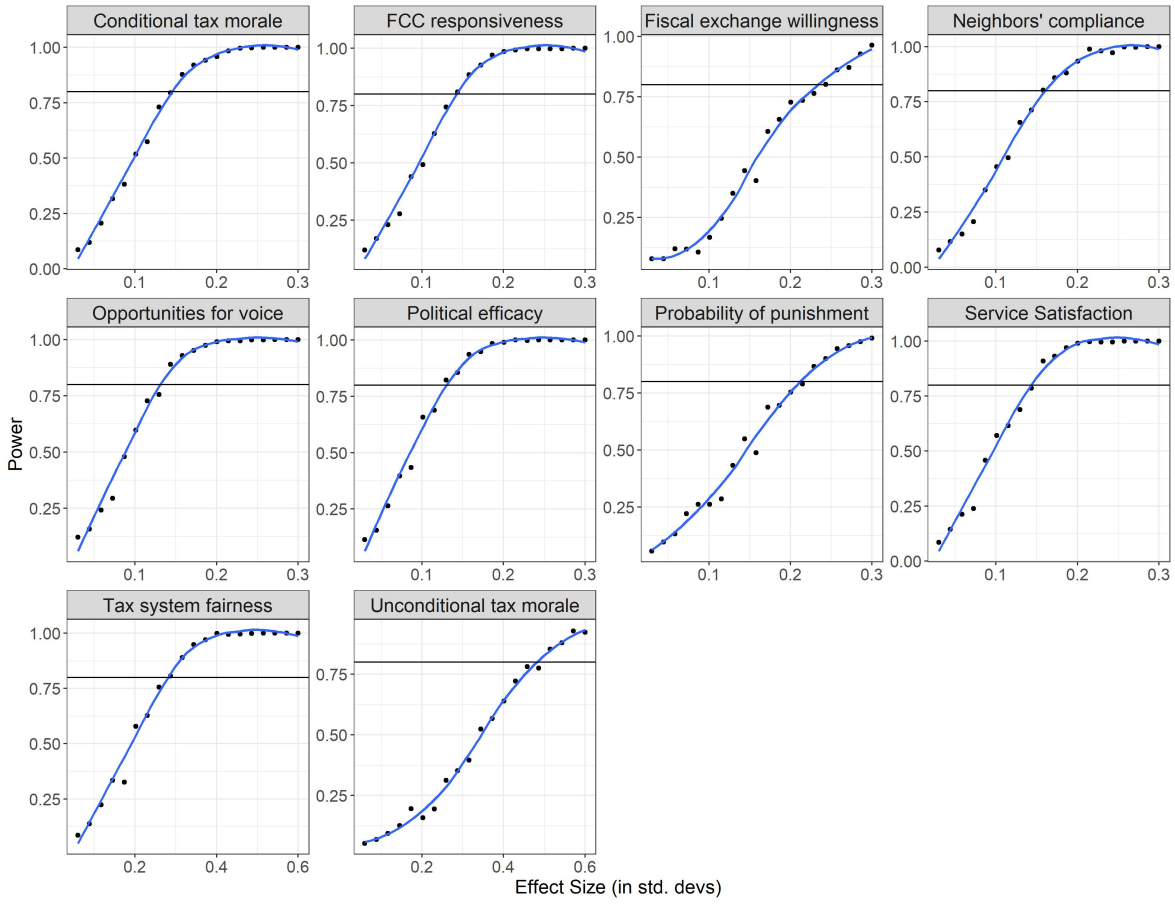
The plots in figure 4 provide results from power simulations for outcome and mechanism variables used in our ITT analysis. Points plot estimated power (y-axis) for a given effect size (x-axis). Effect sizes are given in endline standard deviations of the outcome. For example, the outcome *conditional tax*

[31] These items are likely to overestimate the potential for and occurrence of spillovers, as treated individuals may follow social desirability concerns by exaggerating communications and networks.

morale achieves 80% power at an effect size of just under .15 standard deviations. Given the (endline) standard deviation of 1.56 for this outcome, an effect size of .15 standard deviations is equivalent to a raw effect size of .234 on the 5-point Likert scale.

Most of our outcomes of interest achieve 80% power with effects sizes at or less than .15 standard deviations. The outcomes "tax system fairness" and "probability of punishment for non compliance" have MDE sizes of between .2 and .25 standard deviations; we elect to maintain these outcomes. Unconditional tax morale is underpowered, achieving 80% power at effect sizes of about .5 standard deviations. Given that the endline (control) standard deviation for this outcome is 1.3, this implies a raw effect size of .65 points on the 5-point Likert scale. This is a clear case of ceiling effects hindering our chances of detecting an effect— the endline (control) mean of the outcome is 4.36 out of 5. Given its importance to the tax compliance literature, we elect to maintain this measure in our analysis, but note that we are underpowered to detect small to medium size effects.

Figure 4: Power Simulation (ITT)



11.2 Power Simulation: Heterogeneous Effects Analysis

Figures 5 and 6 provide results from power simulations for heterogeneous effects analyses, where the dependent variable in the analysis is conditional tax compliance and tax compliance behavior, respectively. In each figure, individual plots present power estimates for a given moderating variable. The value on the y-axis should be interpreted as the difference between the effect at the two levels of the moderating variable. For example, when conditional tax morale is the dependent variable (figure

5) and *Age* is the moderating variable, we are powered to detect differences in effect sizes of about .32 standard deviations. So if the true effect for respondents of above median age is zero, and the true effect for respondents of below median age is .32 SD, we can detect this difference with 80% probability.

When examining variation in treatment effects on conditional tax compliance, we can detect differences of about .32 SD when the moderating variable is age, tax liability, education, or political interest. For the remaining five prespecified moderating variables we require differences in effect sizes of about .5 standard deviations to achieve 80% power.

When we turn to variations in treatment effects tax compliance behavior, effects sizes are in percentage points, rather than standard deviations. As the analysis makes clear, for these pre-specified moderating variables, we are only well powered to detect rather large variations in effect size. Age, education, and tax liability require differences in effect sizes of nearly 15 percentage points to achieve 80% power. The remaining six moderating variables need differences in effect sizes of roughly 20 percentage points.

Figure 5: Power Simulation (Y = Conditional Tax Morale

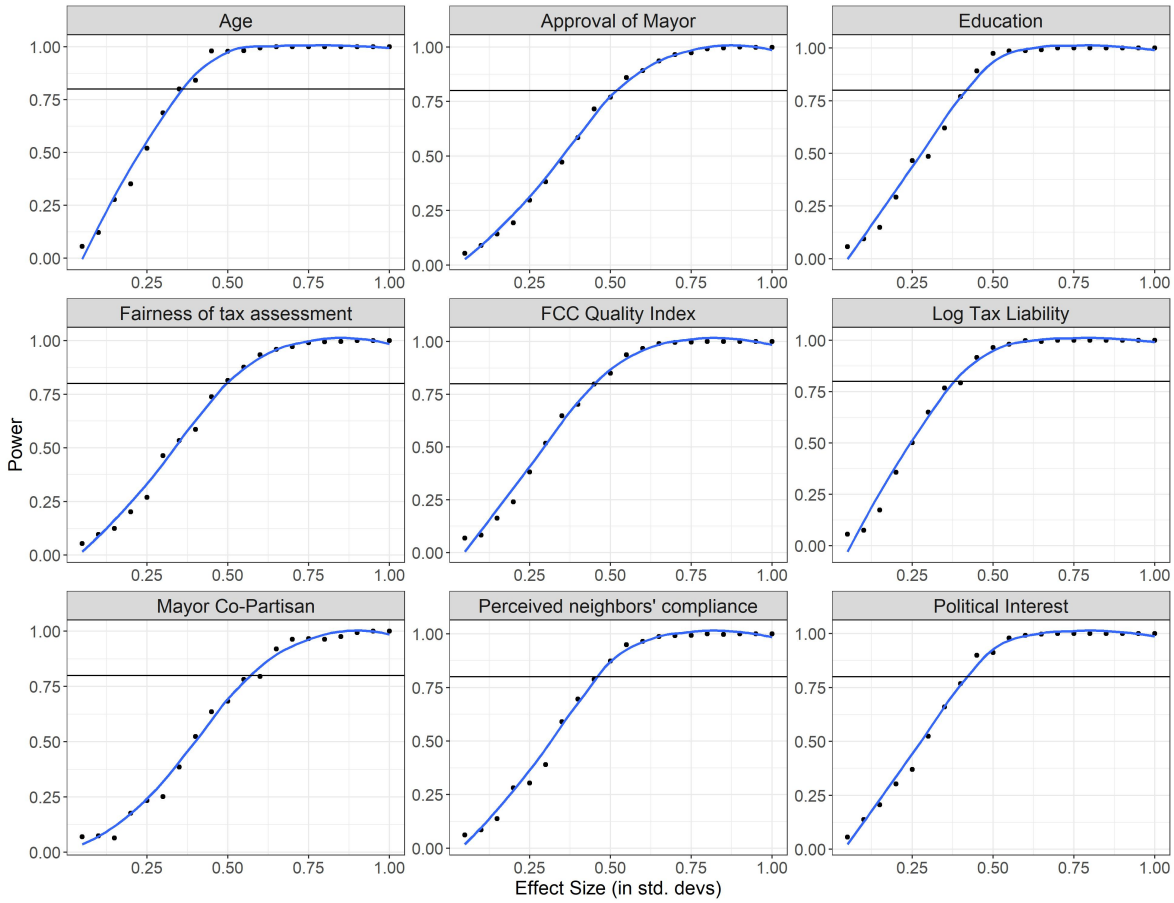
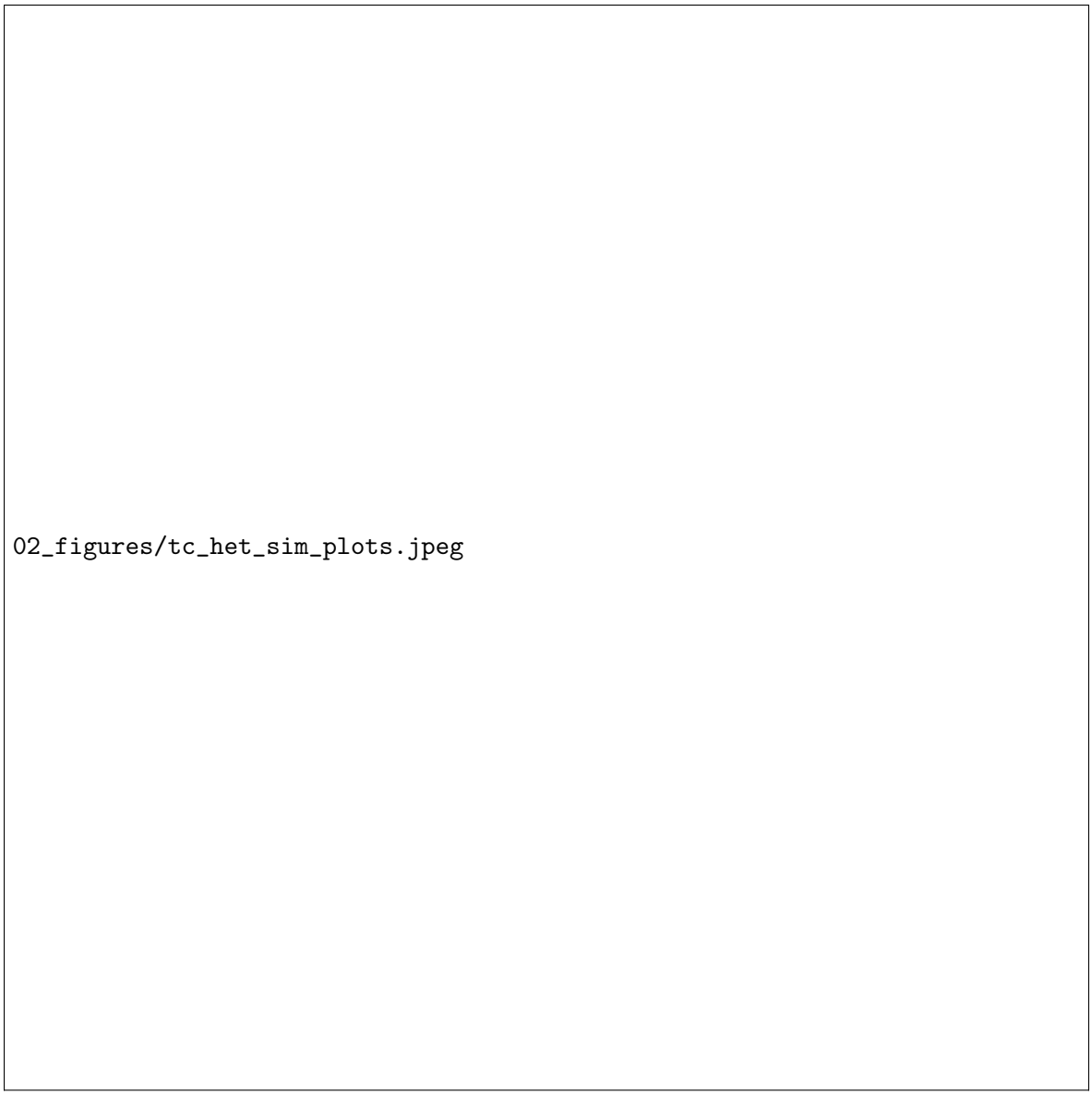


Figure 6: Power Simulation (Y = Unconditional Tax Morale)



12 Research Ethics

We find it important to reflect ethically on several dimensions of this project.

First, much of this project was carried out during COVID-19. Therefore, we took several steps to minimize in-person contact and the risks associated with that contact. Most fundamentally, we shifted the project's primary intervention—the town hall—to an online platform, after having originally conceptualized the intervention as a set of in-person town halls. In addition, we conducted data collection through phone interviews, rather than in-person interviews. Phone surveying followed guidelines from the International Growth Centre for conducting research during the pandemic. For example, while our enumeration team met in person to conduct phone interviews, they followed social distancing and sanitation protocols. Finally, all members of our enumeration team received COVID-19 vaccines. We weighed the risks to our enumeration team against the costs associated with calling the project off. Through conversations with research assistants and project supervisors, we believed that much of our enumeration team would be without a paying job during the pandemic if the project was cancelled. We reasoned that the costs to enumerators of cancelling the project outweighed the risks associated with continuing the project.[32]

Second, only a subset of property owners were eligible to take part in the intervention. We believe that valid equity concerns can be raised about the fact that Freetown citizens who do not own property were not eligible to participate in a participatory budgeting intervention. In addition, eligibility was restricted to property owners (i) with WhatsApp and (ii) with a property above median property value. These latter restrictions were for practical reasons. As the original intervention was originally planned for 2021, and the outcome of interest would be tax compliance in that year, we could only focus on the subset of property owners who received an RDN in 2020.[33] We believe that restricting the intervention to property owners is justified by the scientific goal of the study and because we believe the project has increased the likelihood that all residents of Freetown have a chance to participate in future participatory budgeting programs. Scientifically, we are primarily interested in the relationship between participation in DTH and property tax compliance. Given budget constraints, including non property owners in the intervention would weaken our ability to learn about the effect of the DTH on tax compliance. Moreover, future iterations of the DTHs, to which the Mayor of Freetown has publicly committed, promise to be less restrictive. Freetown residents who were not eligible for this iteration of the DTH are now more likely to be eligible for future participatory budgeting programs, compared to if this DTH project had never taken place. Finally, we do not believe it to be the case that the selected public services only benefit, or even are more likely to benefit, property owners. For example, community water pumps or street taps benefit everyone in the community, not just home owners. Third, we purposefully did not inform participants that the funds for selected services came from donors. While we were generally ambiguous about the source of the funding, in at least one instance, scripted messages from moderators to participants in DTH referred to DTHs as a way to decide on the allocation of some of "FCC's budget". Placards placed at the site of completed projects list the Freetown City Council as the sole implementing partner and the FCC's logo is the only logo on these placards. We believe this deception to be justified by the scientific benefits of the project. While external donors often play a significant role in bankrolling poor local governments, our goal is to study the fiscal contract between government and citizens. We reasoned that acknowledging the external source of funding would make our results more difficult to interpret. Finally, we note that we are not aware of evidence showing that donor credit claiming for donor-funded projects leads to positive outcomes for citizens; in the absence of such evidence, we follow our instinct that donor credit

[32] When making our decision to continue with the project during COVID-19, our research team primarily considered the risks and benefits to our enumeration team. However, we can also point out the additional project benefit of delivering key services (totalling over $45,000) in Freetown.

[33] As described in section 6, as part of a COVID-19 policy to reduce tax burdens on lower income households, only property owners in the top half of the assessment distribution received RDNs in 2020.

claiming for our project is not an *ex ante* normatively superior decision.

References

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American statistical Association*, 103(484):1481–1495.

Barabas, J. (2004). How deliberation affects policy opinions. *American Political Science Review*, pages 687–701.

Bates, R. H. and Lien, D.-H. (1985). A note on taxation, development, and representative government. *Politics & Society*, 14(1):53–70.

Blattman, C., Green, D. P., Ortega, D., and Tobón, S. (2021). Place-based interventions at scale: The direct and spillover effects of policing and city services on crime. *Journal of the European Economic Association*, 19(4):2022–2051.

Boning, W. C., Guyton, J., Hodge, R. H., Slemrod, J., Troiano, U., et al. (2018). Heard it through the grapevine: direct and network effects of a tax enforcement field experiment. Technical report, National Bureau of Economic Research.

Boulianne, S. (2019). Building faith in democracy: Deliberative events, political trust and efficacy. *Political Studies*, 67(1):4–30.

Casey, K. (2018). Radical decentralization: does community-driven development work? *Annual Review of Economics*, 10:139–163.

Chen, J., Humphreys, M., and Modi, V. (2010). Technology diffusion and social networks: Evidence from a field experiment in uganda. *Manuscript, Columbia University*.

Collier, P. and Vicente, P. C. (2014). Votes and violence: Evidence from a field experiment in nigeria. *The Economic Journal*, 124(574):F327–F355.

Drago, F., Mengel, F., and Traxler, C. (2015). Compliance behavior in networks: Evidence from a field experiment.

Esterling, K. M., Neblo, M. A., and Lazer, D. M. (2011). Means, motive, and opportunity in becoming informed about politics: A deliberative field experiment with members of congress and their constituents. *Public Opinion Quarterly*, 75(3):483–503.

Farrar, C., Fishkin, J. S., Green, D. P., List, C., Luskin, R. C., and Paluck, E. L. (2010). Disaggregating deliberation's effects: An experiment within a deliberative poll. *British Journal of Political Science*, pages 333–347.

Friess, D. and Eilders, C. (2015). A systematic review of online deliberation research. *Policy & Internet*, 7(3):319–339.

Fujiwara, T. and Wantchekon, L. (2013). Can informed public deliberation overcome clientelism? experimental evidence from benin. *American Economic Journal: Applied Economics*, 5(4):241–55.

Fukumoto, K. (2015). Missing data under the matched-pair design: a practical guide. Technical Report, Presented at the 32nd Annual Summer Meeting of Society for ....

Gerber, A. S. and Green, D. P. (2012). *Field experiments: Design, analysis, and interpretation*. WW Norton.

Gonçalves, S. (2014). The effects of participatory budgeting on municipal expenditures and infant mortality in brazil. *World Development*, 53:94–110.

Habermas, J. (1975). *Legitimation crisis*, volume 519. Beacon Press.

Jaidka, K., Zhou, A., and Lelkes, Y. (2019). Brevity is the soul of twitter: The constraint affordance and political discussion. *Journal of Communication*, 69(4):345–372.

Kies, R. (2010). *Promises and Limits of Web-deliberation*. Springer.

King, G., Gakidou, E., Ravishankar, N., Moore, R. T., Lakin, J., Vargas, M., Téllez-Rojo, M. M., Hernández Ávila, J. E., Ávila, M. H., and Llamas, H. H. (2007). A "politically robust" experimental design for public policy evaluation, with application to the mexican universal health insurance program. *Journal of Policy Analysis and Management*, 26(3):479–506.

Levi, M. (1988). *Of rule and revenue*. University of California Press.

López-Moctezuma, G., Wantchekon, L., Rubenson, D., Fujiwara, T., and Pe Lero, C. (2022). Policy deliberation and voter persuasion: Experimental evidence from an election in the philippines. *American Journal of Political Science*, 66(1):59–74.

Luskin, R. C., O'Flynn, I., Fishkin, J. S., and Russell, D. (2014). Deliberating across deep divides. *Political Studies*, 62(1):116–135.

Mansuri, G. and Rao, V. (2013). Localizing development: Does participation work?

Martin, L. (2014). Taxation, loss aversion, and accountability: theory and experimental evidence for taxation's effect on citizen behavior. *Unpublished paper, Yale University, New Haven, CT*.

Meiselman, B. S. (2018). Ghostbusting in detroit: Evidence on nonfilers from a controlled field experiment. *Journal of Public Economics*, 158:180–193.

Miguel, E. and Kremer, M. (2004). Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica*, 72(1):159–217.

Neblo, M. A., Esterling, K. M., Kennedy, R. P., Lazer, D. M., and Sokhey, A. E. (2010). Who wants to deliberate—and why? *American Political Science Review*, pages 566–583.

North, D. C. and Weingast, B. R. (1989). Constitutions and commitment: the evolution of institutions governing public choice in seventeenth-century england. *Journal of economic history*, pages 803–832.

Parthasarathy, R., Rao, V., and Palaniswamy, N. (2019). Deliberative democracy in an unequal world: A text-as-data study of south india's village assemblies. *American Political Science Review*, 113(3):623–640.

Prichard, W. (2015). *Taxation, responsiveness and accountability in Sub-Saharan Africa: the dynamics of tax bargaining*. Cambridge University Press.

Prichard, W., Custers, A., Dom, R., Davenport, S., and Roscitt, M. (2019). *Innovations in tax compliance: Conceptual framework*. The World Bank.

Ross, M. L. (2004). Does taxation lead to representation? *British journal of political science*, pages 229–249.

Sandefur, J., Birdsall, N., Fishkin, J., and Moyo, M. (2020). Deliberative democracy and the resource curse: A nationwide experiment in tanzania.

Sexton, R. (2017). The unintended effects of bottom-up accountability: Evidence from a field experiment in peru. Technical report, Working Paper.

Sheely, R. (2015). Mobilization, participatory planning institutions, and elite capture: Evidence from a field experiment in rural kenya. *World Development*, 67:251–266.

Sinclair, B., McConnell, M., and Green, D. P. (2012). Detecting spillover effects: Design and analysis of multilevel experiments. *American Journal of Political Science*, 56(4):1055–1069.

Sjoberg, F. M., Mellon, J., Peixoto, T. C., Hemker, J. Z., and Tsai, L. L. (2019). Voice and punishment: A global survey experiment on tax morale. *World Bank Policy Research Working Paper*, (8855).

Tilly, C. (1985). War making and state making as organized crime. *Violence: A reader*, pages 35–60.

Torgler, B. (2005). Tax morale and direct democracy. *European Journal of Political Economy*, 21(2):525–531.

Touchton, M. R., Wampler, B., and Peixoto, T. C. (2019). Of governance and revenue: Participatory institutions and tax compliance in brazil. *World Bank Policy Research Working Paper*, (8797).

Weigel, J. L. (2020). The participation dividend of taxation: How citizens in congo engage more with the state when it tries to tax them. *The Quarterly Journal of Economics*, 135(4):1849–1903.

Appendices

A Outcome Summary Statistics

	Mean	SD	Min	Q25	Q50	Q75	Max	N	Missing
Unconditional tax morale	3.77	1.55	1	2	5	5	5	1797	12
Service conditional tax morale	1.96	0.96	1	1	2	3	3	1805	4
Fiscal exchange willingess	4.19	1.22	1	4	5	5	5	1805	4
Satisfaction with FCC service provision	3.64	1.17	1	3	4	4	5	1790	19
Opportunities for voice	2.12	1.00	1	1	2	3	4	1719	90
Political efficacy	1.76	1.14	1	1	1	2	5	1794	15
FCC responsiveness	3.17	1.18	1	2	4	4	5	1712	97
Tax system fairness	2.12	0.79	1	1	2	3	3	1112	697
Neighbors' compliance	5.13	2.41	0	3	5	7	10	1138	671
Percieved probability of punishment	4.06	1.11	1	4	4	5	5	1788	21

Table 6: Summary statistic for outcome variables at baseline (control group)

	Mean	SD	Min	Q25	Q50	Q75	Max	N	Missing
Unconditional tax morale	3.94	1.53	1	2	5	5	5	1645	164
Service conditional tax morale	1.72	0.93	1	1	1	3	3	1647	162
Fiscal exchange willingess	4.00	1.25	1	4	4	5	5	1648	161
Satisfaction with FCC service provision	3.61	1.06	1	3	4	4	5	1648	161
Opportunities for voice	2.33	0.92	1	2	2	3	4	1636	173
Political efficacy	1.62	1.02	1	1	1	2	5	1646	163
FCC responsiveness	3.36	1.06	1	2	4	4	5	1623	186
Tax system fairness	2.15	0.69	1	2	2	3	3	1147	662
Neighbors' compliance	5.97	2.29	0	5	6	8	10	1498	311
Percieved probability of punishment	4.24	0.98	1	4	5	5	5	1648	161

Table 7: Summary statistic for outcome variables at Midline (control group)

	Mean	SD	Min	Q25	Q50	Q75	Max	N	Missing
Unconditional tax morale	4.36	1.30	1	5	5.00	5	5	1397	412
Service conditional tax morale	2.52	1.56	1	1	2.00	4	5	1402	407
Fiscal exchange willingess	4.03	1.29	1	4	4.67	5	5	1402	407
Satisfaction with FCC service provision	3.47	1.21	1	2	4.00	4	5	1400	409
Opportunities for voice	2.16	0.92	1	1	2.00	3	4	1392	417
Political efficacy	1.63	1.02	1	1	1.00	2	5	1395	414
FCC responsiveness	3.31	1.13	1	2	4.00	4	5	1380	429
Tax system fairness	2.38	0.78	1	2	3.00	3	3	1394	415
Neighbors' compliance	5.92	2.45	0	4	6.00	8	10	1210	599
Percieved probability of punishment	4.14	1.04	1	4	4.00	5	5	1395	414

Table 8: Summary statistic for outcome variables at Endline (control group)

B Descriptive hypotheses

Below, we lay out two sets of descriptive analyses.

B.1 Descriptive Analysis 1: Who Participates Actively?

In line with the research on the participation gap in participatory settings, we expect the following individual characteristics to predict the frequency of participation, which we measure by the number of overall number of contributions in a DTH. We hypothesize that:

D1a: Men participate more than women.
D1b: Property owners with higher levels of education participate more.
D1c: Wealthier, higher income earning property owners participate more.
D1d: Property owners with higher levels of political interest participate more.
D1e: Property owners who perceive the new property tax system to be unfair participate more.
D1f: Co-partisans of the Mayor (APC supporters) are more likely to participate.

In addition, we expect that following group level characteristic impacts participation:

D1g: Overall DTH participation is higher in DTH groups with more ethnic homogeneity.

B.2 Descriptive analysis 2: Who Sets the Agenda? Who Influences Eventual Choices over Services?

Agenda setting in a deliberative context occurs when a participant introduces a topic—e.g., a preferred service—which is then being referenced or replied to by other participants. A participant is influential when they can persuade other participants to align preferred services with their own. We make several hypotheses about who sets the agenda and is influential.

D2a: Men are better able to shape the discussion and influence service choice.
D2b: Property owners with higher levels of education are better able to shape the discussion and influence service choice.
D2c: Wealthier, higher income earning property owners are better able to shape the discussion and influence service choice.
D2d: Property owners with higher levels of political interest are better able to shape the discussion and influence service choice.
D2e: Co-partisans of the Mayor (APC supporters) are better able to shape the discussion and influence service choice.
D2f: Participants with higher tax liability are better able to shape the discussion and influence service choice.

C Distance to Closest Study Property

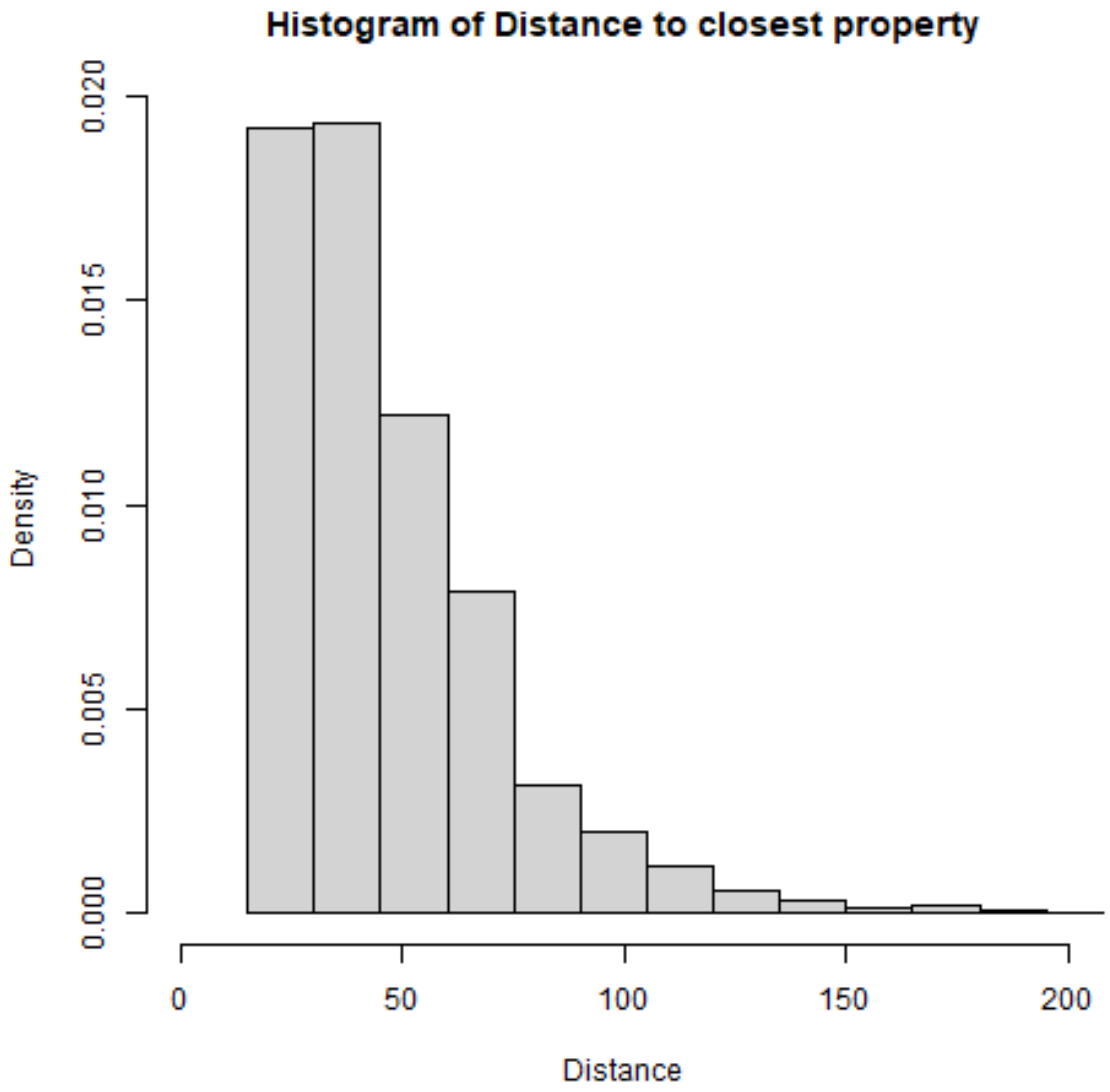


Figure 7: Histogram of minimum distance (in meters) between study properties

D Covariate Summary Statistics

Table 9: Summary statistic for control variables

	Mean	**SD**	**Min**	**Max**	**N**
Log total tax liability 2021 (USD)	3.75	1.15	0	8.38	3618
Total number of property with any liability in 2021	1.91	1.47	0	18.00	3618
Tax Compliance (2021)	0.29	0.45	0	1.00	3618
Water	0.47	0.50	0	1.00	3618
Drainage	0.36	0.48	0	1.00	3618
In informal settlement	0.06	0.24	0	1.00	3618
Fencing and gate	0.39	0.49	0	1.00	3618
Street status: Footpath	0.30	0.46	0	1.00	3618
Street status: Major	0.08	0.27	0	1.00	3618
Street Status: Minor	0.63	0.48	0	1.00	3618
Street condition: Average	0.29	0.45	0	1.00	3618
Street condition: Like New	0.16	0.36	0	1.00	3618
Street condition: Poor	0.55	0.50	0	1.00	3618
Ease of street access: Difficult	0.50	0.50	0	1.00	3618
Ease of street access: Easy	0.50	0.50	0	1.00	3618
Window quality: Not Traditional	0.45	0.50	0	1.00	3618
Window quality: Traditional	0.55	0.50	0	1.00	3618
Garage: No	0.93	0.26	0	1.00	3618
Garage: Yes	0.07	0.26	0	1.00	3618
RDN Type 2021: 1	0.17	0.37	0	1.00	3618
RDN Type 2021: 2	0.16	0.37	0	1.00	3618
RDN Type 2021: 3	0.17	0.37	0	1.00	3618
RDN Type 2021: 4	0.17	0.37	0	1.00	3618
RDN Type 2021: 5	0.17	0.38	0	1.00	3618
RDN Type 2021: 6	0.17	0.38	0	1.00	3618

E Project Pictures

Figure 8: Project implemented in Ward 418.

Figure 9: Project implemented in Ward 442.

Figure 10: Project implemented in Ward 444.

F Number of Units in Spillover Estimation

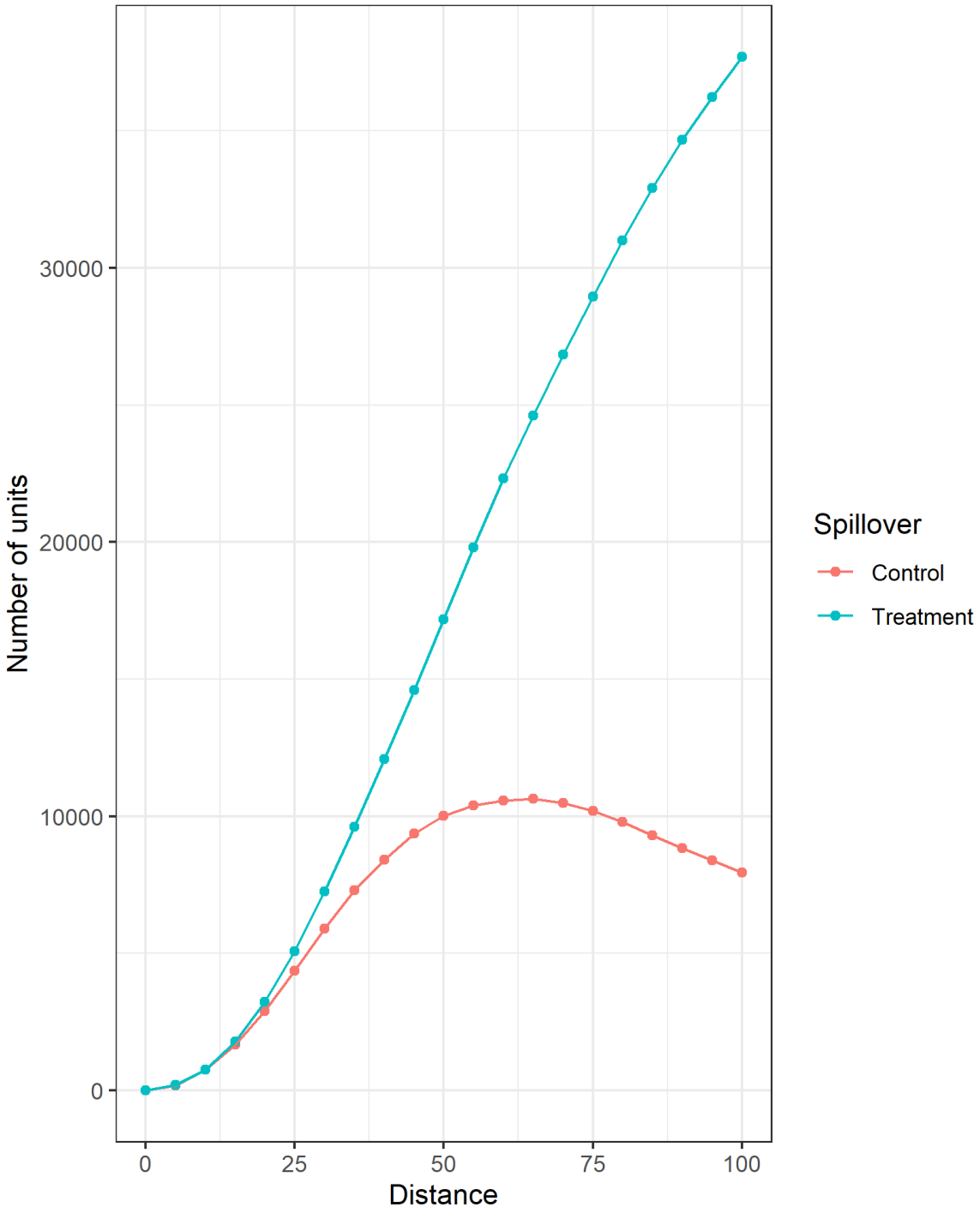
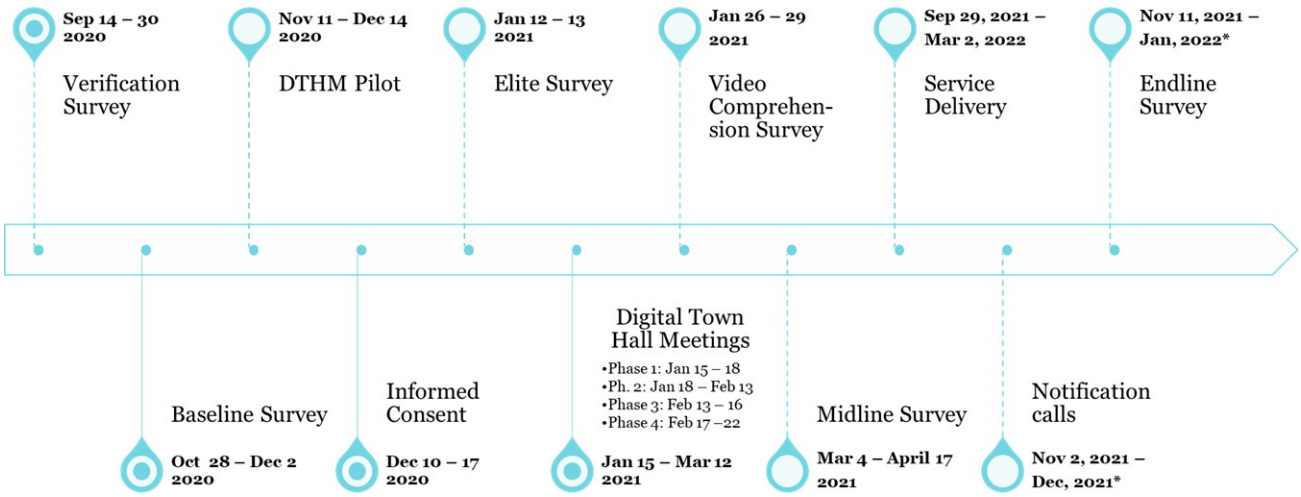


Figure 11: Units in each spillover condition at different values of D

G Manuals and Instructions

G.1 Timeline

Figure 12: Intervention Timeline



*Note that notification calls and endline surveys in one ward, Tengbeh Town, were delayed by two months due to contractual issues with the construction firm.

DIGITAL TOWN HALL MEETINGS

A Technical Manual for Enumerators

January 2021

GLOSSARY

Participant – In the context of the DTHM, a **participant** is one of the people you talked to in the informed consent survey. These people are the ones you will be managing when you 'mediate' the group chats, and they will be the only ones allowed to discuss and vote on development policy in the ward they live in.

Moderator – This is **you**. You are responsible for keeping the DTHM running, and making sure that participants successfully participate in it!

Mediating – "**Managing people**". This is about making sure that the participants in the DTHM are comfortable and prepared to participate. This might mean preventing arguments or fights, answering questions that get sent to you, and posting important content during each phase.

Add any other words, terms, or phrases which you think you'll need to remember.

INTRODUCTION

Last time, we covered the basic idea of a Digital Town Hall Meeting (DTHM). By now, you should feel comfortable with the structure of the DTHMs, and understand what the participants will be required to do as part of the process. It is very important that you do, since participants will need guidance from you throughout the DTHM. Of course, we will only launch the DTHMs when you are confident managing them, and Michael and Emile will continue to support you throughout the entire process. If you ever need to ask questions – and it is likely that you will, since this is a very unusual project – please do not hesitate to ask these to either Michael or Emile.

What are the four phases of the DTHM?

What we did last time was to first understand the DTHMs from the perspective of the **participant**. This means things like learning the phases of the DTHM, what sort of messages are expected in the chat, and how the winning development project gets selected. But the next step for us is to consider what **you** will need to do in order to make this possible. This technical manual goes into more depth on exactly what WhatsApp tools we expect you to use, and how you can use them effectively to create a successful DTHM.

YOUR ROLE

YOUR TWO JOBS

We also spoke a little bit about the two halves of your job – **mediating**, and **synthesising**. You will need to be able to do both of these, so that we can adapt to the challenges of the DTHM.

MEDIATING

A good way to understand mediating is '**managing people'**. This includes things like:

- Preventing and defusing arguments between DTHM participants
- Answering people's questions
- Posting the mayor's videos at the start of each phase

There are various different skills that you will need to be able to fulfil these tasks. For example, to be able to explain to participants how the DTHM works, you yourself will need to feel confident in understanding the DTHM process.

What other skills do you think you will need for mediating?

SYNTHESISING

Synthesising is about writing a summary of the **information in the group chat**.

Think about a WhatsApp group chat that you regularly post in. When someone posts a funny, interesting, or provocative message in this group chat, the chances are that lots of the members of the chat join in to discuss. Very quickly, tens, maybe even hundreds of messages build up in the chat, making it difficult to scroll back to the start to see what caused the discussion. The chances are that the group chat you're thinking of has a lot less than 66 people in it!

It's therefore not an easy task for you to read all of these messages. If it's difficult for you, then it will be impossible for the mayor, who may only have a few minutes to read. So that the mayor and ward councillors can do their job, it will be up to you to:

1. Read through the chat
2. Summarise the content of the chat

What will you need to look for when reading through the chat?

If you do this well, then it will be much easier for the mayor and ward councillors to come up with a good policy menu that fits people's needs. This is very important for the DTHM to be a success. Don't forget that, for the mayor and ward councillors, the feedback that the DTHM provides, and the summary that you create, may be very helpful for other projects beyond the DTHM.

What do you think might happen if you write a bad summary?

Why are we asking you to do both of these things? Why not give one of you the task of mediating, and the other the task of synthesising?

THE TEAM

You have already been told about the team on this project – there are:

- 10 moderators
    - Split into 5 teams of 2 people
- 3 supervisors
    - Inkia
    - AKD
    - ABJ
- 2 RAs
    - Michael
    - Emile
- A handful of principal researchers (who are all outside Salone)

All of us on this team are working **together**. Everyone will play a different role, but every single one of us will contribute to making these DTHMs a success.

Why are you organised into teams? (The RAs will tell you who you're working with!)

Working with the RAs

One of the great things about this project is that, because it is so important and unusual, the RAs will be very involved. From start to finish, we'll be on hand to work alongside you, and we'll be interested to see what is working well for you, and what you're finding difficult. Rest assured that this is not because we are unhappy with your work. This is because there may be certain parts of the DTHM that can be designed better in future, or because you identify a problem that other moderators might need to know about.

We encourage you to ask questions and communicate with us frequently, so that we can help you, and understand these things. In fact, **at the end of each working day, we join together to have a 20-minute conversation about how that day has gone for you**. But if you have a question you need to ask, the RAs will always be at the venue with you, or contactable on the phone. They will always be happy to take your questions as soon as they occur to you.

WhatsApp

As moderators, **you will be the administrators of the WhatsApp groups**. The RAs and your supervisor will also join the DTHMs as an administrator, so that if an emergency occurs, they can help. **Nobody else may be made an administrator of the groups**. You are working in a highly political environment, where protecting the participants is your number one priority. This is the first, and one of the most important ways that you can do this.

What sorts of things might someone be able to do if they are wrongly made an administrator?

We will return to the topic of how we will do each specific phase in a little while. First, there are some tools, and problems, which you will need to be aware of throughout the whole DTHM. Let's recap:

Problems

Can't Mute Individuals

If an individual starts to post inappropriate content to the WhatsApp group, there is no way to temporarily mute only this person.

Depending on the situation, alternatives might include:

- Muting everyone except administrators
- Kicking the participant

- Telling the individual to stop, via a:
    - Private message
    - Public message (in the group)

Think about what kind of inappropriate behaviour might need different actions. Would you ever use a combination of the alternatives above?

Can't Delete Participants' Messages

If someone *does* post inappropriate content to the WhatsApp group, the first priority is to prevent them from doing it again. But once we have achieved this, we cannot delete the offending message. This makes it doubly important that we **prevent** inappropriate behaviour as much as we can – because it's difficult to react afterwards!

Depending on the situation, we might:

- Ask the person to delete their message
- 'Cover up' the message by posting ourselves

Can't Hide Participants' Details

As soon as a participant joins the WhatsApp group, their profile will be there for all to see. That includes:

- Profile photo
- Name
- Mobile number

For this, we have to warn participants ahead of time, and reassure participants that they are **safe** in the DTHM, reminding them that the DTHM is a safe space to share opinions. We did this when you read the 'informed consent' message in the previous survey.

DISCUSS: Should we message participants individually to warn them before the WhatsApp group starts?

No Polling Feature

In Phase Four, participants will need to vote on their preferred policies. Ideally, we would do this in WhatsApp, to make it **as easy as possible.** But WhatsApp cannot do this on its own. Instead, we will need to **post a link to another website** to get people to vote.

DISCUSS: Can you think of any other problems we may face? Let's work out together how we can address them.

Tools

It's not all bad, however. WhatsApp also has a wide variety of tools or functions which will make our lives much easier. We'll need to be aware of these and use these appropriately throughout the DTHM.

Only-Administrators Posting

As an administrator of the WhatsApp group, you will be able to restrict posting to only administrators (us!).

DISCUSS: When is it appropriate to do this?

Kicking Participants

Administrators are also able to kick other group chat members at any time.

DISCUSS: When is it appropriate to do this?

Starred Messages

Anyone in the chat is able to 'star' messages, which means that they can be saved and read in a separate tab, personal to them.

DISCUSS: Why is this useful/when is it appropriate to do this?

You have a good idea now of what your job will be in the DTHM. Can you think of any other good tools that would help you?

GUIDELINES

We need to make rules for the chat, and guidelines for *ourselves*: when do we take serious action, and when do we need to intervene?