Pre-Analysis Plan
"Can you hear me now?": Experimental evidence on improving public service delivery through non-electoral citizen participation

Saher Asad[*]   Asim I. Khwaja[†]   Tiffany M. Simon[‡]

July 10, 2025

1   Introduction

Improving public service delivery can support economic and social development and strengthen citizen-government linkages, particularly in developing country contexts. Yet, in cases where service delivery is inadequate and citizens perceive and/or experience barriers to engagement with the government, what effective means do citizens have to hold the government representatives accountable? In developing countries, such as Pakistan, policy actors often are eager to engage with the public as can be seen by the growth in various government portals and complaint hotlines. However, the take-up of these mechanisms to reach policy actors is low and they are often too complicated for citizens to engage with at an individual level. Our project uses public education, an issue that citizens deeply care about in Pakistan, as a case in point, and we use community meetings as a means to encourage citizen interaction with policy actors to examine how these interactions can impact policy outcomes. Specifically, we use a randomized control trial that introduces community-based mobilization interventions to improve public schooling in Pakistan. Based on pilot work, we vary these interventions by: (i) policy actor type – whether citizens approach a bureaucrat directly or exert pressure through a political route and (ii) citizen gender – whether the citizens participating are women or men. In addition, for each we also include a variation in which citizens' interaction with the policy actor is more directly supported and facilitated by an NGO. We examine impact of the intervention on citizen political action, problem resolution, and school quality. Please refer to the AEA RCT Registry (AEARCTR-0011478) for full details of the intervention and study design.

[*]World Bank, email: sasad1@worldbank.org
[†]Harvard Kennedy School, email: khwaja@hks.harvard.edu
[‡]Princeton University Politics Department, email: tsimon@princeton.edu

2   Study Design

Our original study design split our original sample into study into two phases – "Phase I," in which we would run our study in 48 of an original 480 sample villages in the district of Khushab and use qualitative learnings to refine logistics, intervention delivery, survey instruments, and sampling design, and then implement these changes in an updated PAP prior to roll out of the remaining sample of 432 villages in "Phase II."

While Phase I was successfully launched and completed by summer 2023, the rollout of Phase II faced repeated delays. The political instability in Pakistan during 2023 and early 2024 – including the arrest and sentencing of former Prime Minister Imran Khan, the dissolution of the National Assembly, and the postponement of parliamentary elections – created significant challenges for study implementation. Public sentiment was politically charged, and the absence of elected assembly members diminished the relevance of our treatment, which relies in part on providing information about political representation and the contact information and encouragement to contact a treated participant's elected Member of the Provincial Assembly.

Compounding this, many bureaucratic appointments remained vacant during this period, limiting the government's ability to respond to public needs – a second essential component for our intervention's effectiveness. Restrictions on public gatherings due to protests, along with bans on data collection in the weeks surrounding the election further delayed field activities. These delays also led to previously secured government approvals for fieldwork to lapse, requiring several months to renew.

Following the election, the project experienced substantial staff turnover, necessitating the rehiring and retraining of staff and field teams across multiple districts. The timing of Ramzan and school summer holidays further delayed planned fieldwork until after summer 2024.

Given the nearly two-year gap between Phase I and Phase II – and the major changes in political and administrative context – we have decided to exclude Phase I data from our analysis. While we attempted to expand our sample in Phase II districts to compensate, we effectively max out our sample in these areas given our sampling strategy (all villages with a government primary school and a government primary school) and the announcement of the Public Schools Reorganization Program in summer of 2024 (further explained in the section below), which led to a loss of an additional 16 eligible villages from our sample. However, power analyses conducted prior to the decision to drop Khushab data from the study found that this reduction in sample size had a minimal impact on the minimal detectable effect size of key outcomes.

This pre-analysis plan applies to both midline and endline analyses. Based on findings from the midline, we may revise our endline survey to include additional questions. Based on midline results, we may also delay the endline if it's warranted to conduct a refresher of the treatment for treated villages/individuals.

2.1   Sample

We sourced our sample from the Punjab Education Department's School Census, which contains a comprehensive list of all government schools in Punjab. We then merged this

data with information on National Rural Support Program (NRSP) Village Organizations provided by NRSP, allowing us to identify only those villages where NRSP has been active that have a primary government school.

From this list, we excluded schools that were either already selected or expected to be selected under the Public School Restructuring Program (PSRP). Information about these schools was obtained from the official PSRP Terms of Reference document[1] and verified through coordination with the National Rural Support Program (NRSP).

We excluded these schools from our sample for two main reasons. First, under the PSRP, responsibility for school performance is expected to shift to private organizations, with the government playing only a monitoring role. Since our intervention focuses on encouraging parents to engage with government policy actors, targeting schools managed by private entities would fall outside the scope of our project. Second, the program's design and implementation timeline remain unclear. If a school transitions to private management during the course of our study, this could alter the incentives of government actors to respond to parental outreach, reducing both the effectiveness of our intervention and the insights we could draw. Even if the conversion has not occurred by the start of the school year, policymakers may still disregard parental concerns, anticipating that the performance of these schools will soon fall outside their realm of responsibility.

Our full study sample consists of 433 randomly selected villages in two districts in which NRSP has the strongest presence, as further detailed below. Our sample is stratified on school gender – in half (216) of the villages, we have selected a girls school as the government primary school of focus, while the other half (217) we will select a boys' school as the government primary school of focus.

2.1.1 Villages

Our target population is the rural poor of Punjab, Pakistan's largest state. Of the 18 districts in Punjab where our implementing partner, NRSP, works, we have selected two districts in which they have the most villages with local support organizations – Bahawalpur and Bahawalnagar.

2.1.2 Households

A key part of our household sampling strategy is to identify the relevant study population. Our main criterion for "relevance" is being the parent or guardian of a child attending the sampled government school. However, we also want to ensure that our sample includes individuals who are likely to be exposed to our treatment (that is, individuals who attend our community meeting intervention). Following learnings from our pilot in Khushab, we have amended our listing as detailed below.

In each village we will survey 10 households randomly selected from a list-frame of 20-25 households whose children attend a selected local government primary school and are invited to participate in our intervention. While only a subset of 10 respondents will be interviewed at baseline and endline, we will invite all listed respondents to the meeting. This means we will effectively have two samples – a treatment sample, made

---

[1] See https://www.pef.edu.pk/pdf/Ad/PSRP/PSRP_TORs.pdf.

up of individuals who are invited to participate in the intervention, and a survey sample, made up of a random subset of the treatment sample, who are surveyed at baseline. This dual sample ensures that our treatment involves an appropriate number of individuals, while not surveying the entire treatment sample reduces our per observation cost, allowing us the budget to maximize our overall village sample. Inviting a larger group for meetings also helps intensity our effects by making sure that those aren't able to join the meetings due to day-to-day issues may be able to learn from those who are. Our power calculations on key outcomes suggest that based on ICC we need 10 individuals per village/cluster to be able to detect sufficient effect sizes. Enumerators follow the following protocol to identify households eligible for the study:

1. Draw a location map

   (a) Enumerators arrive in the village and locate the settlement in which the selected government primary school is located.

   (b) Enumerators then meet with a prominent village resident such as a mosque head, prominent shopkeeper or guard, etc. and with their assistance draw a map of the settlement boundaries, school location, streets with households and main landmarks. They also note the location of up to four closest settlements.

2. Conduct the listing survey

   (a) Enumerators then draw a line through the middle of the map splitting the settlement in half. Then, beginning in the upper right corner of the half of the settlement map that contains the school, enumerators visit every household and complete a short listing survey identifying whether the household contains a government primary school-attending child.

   (b) If enumerators have not identified 25 eligible households after completing the first half of the settlement, they continue into the second half.

   (c) If enumerators survey all households in the settlement but do not find at least 25 eligible households then go to the nearest settlement (as previously identified) and follow the same procedure of mapping and listing in this settlement until 25 eligible households has been achieved for this village.

   (d) If, after following this procedure in all settlements in the village, 25 eligible households are still not found, satellite imagery and school enrollment data are reviewed to check for missed settlements and verify complete exploration of entire village. Enumerators must revisit the village, consult a different local informant, and use their map to identify and cover any missed settlements. If, after revisits, at least 15 households are listed, the listing is considered complete. If there are still less than 15 listed households, the research team verifies that the entirety of the village and nearby settlements have been covered using satellite imagery (Google Earth). If all settlements are exhausted, the protocol was followed, and at least 10 households are listed, listing can stop. If fewer than 10 households are listed, listing is expanded to nearby villages only if they share the same MPA as the original target village.

A household is considered eligible if it meets the following criteria

1. A child enrolled in grades 4 and below in selected government primary school resides there

2. Either mother and father permanently reside in the household (or in the absence of mother or father, a male or female adult (same gender as absent parent) responsible for the education of children permanently resides in the household)

3. The male and female respondent both consent to participate in the survey

4. The male and female respondent both agree to be contacted on phone

5. The male and female respondent both agree to be contacted at their address in future

In terms of respondents, because we want to compare same-gender respondents irrespective of their treatment status, we survey both male and female respondents in 100% of the control villages. In 75% of male (female) treatment villages, we survey only male (female) respondents, and in 25% of male (female) treatment villages, we survey both male (female) and female (male) respondents. This should give us approximately 6,500 respondents (2,900 respondents in control, 3,600 respondents in treatment).

2.2 Randomization

We will randomize treatment – holding of community meetings – at the village level. Within each of our two districts, we identify all villages that have both a government primary school and have ever had an NRSP local support organization. Note that NRSP is broadly active in par'ticular districts in Punjab, so this selection ensures NRSP presence to execute the intervention. Within this set, randomization is stratified by the gender of the school in the village such that we have an equal number of girls and boys schools in our sample. The study covers two districts in Punjab, and villages in these districts will be assigned to each of the treatments as presented in Figure 1.

In total, we will assign 145 villages to pure control and 288 villages to treatment (with 36 villages in each of the eight treatment arms). Based on the power calculations, we are only interested in and powered up to detect effects between control and any 2 variations at a time (e.g. Actor x Gender, Actor x Facilitation or Gender x Facilitation). Treatment will be stratified across all tehsils (district sub-units) in a district and across these 433 villages as follows in Table 1.

Table 1: Treatment Randomization by School Gender

		Treatment Group									Total
		Male-Bureaucrat-Facilitated	Male-Politician-Facilitated	Male-Bureaucrat-Unfacilitated	Male-Politician-Unfacilitated	Female-Bureaucrat-Facilitated	Female-Politician-Facilitated	Female-Bureaucrat-Unfacilitated	Female-Politician-Unfacilitated	Control	
School Gender	Boys' Schools	18	18	19	17	18	18	18	18	73	217
	Girls' Schools	18	18	17	19	18	18	18	18	72	216
	Total	36	36	36	36	36	36	36	36	145	433

Figure 1: Experimental Design
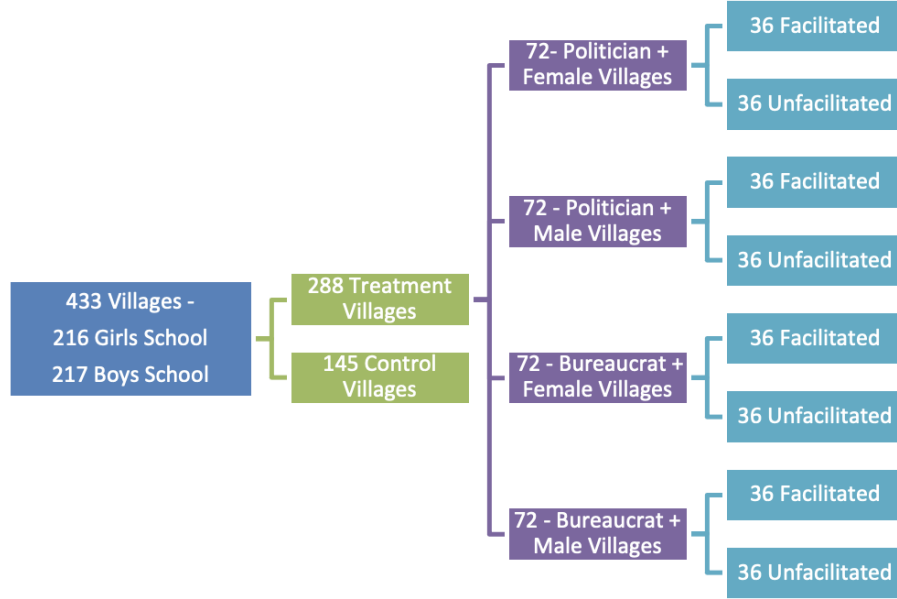
2.3 Data collection

We conduct a brief tracker survey via phone with participants approximately 4 weeks after the final (fourth) meeting in our intervention.

We conduct a full in-person endline survey six months after the final meeting in our intervention. We may delay rollout of the endline survey by 2-4 weeks pending our examination of data from our midline tracker.

3 Outcomes

We will examine impact along three main types of outcomes of interest using both household level surveys and administrative data.

Research questions:

- How does citizen engagement with the government impact public service provision, namely perceptions of public service delivery?

- How does citizen engagement with the government impact perceptions of government effectiveness?

- How does this effect vary by gender and policy actor?

3.1 Primary outcomes

Our main hypothesis is that removing frictions related to coordination, information and self-efficacy will enable citizens who want to improve their children's government primary school to come together and contact policy actors and resolve the most important problems they are facing in their children's schools. In cases where we have more than one variable capturing each of the outcomes, relevant indices will be constructed. To study this hypothesis, we will investigate impacts of our intervention on the following outcomes:

1. Whether citizens take action on their education problem

2. Whether citizen action results in problem resolution

3.2 Other outcomes and channels

We also intend to explore several secondary outcomes which will highlight the different channels through which our treatments may be working via our theory of change:

1. Awareness of education problems. Citizens are often unable to act themselves because they aren't aware of the universe of problems at a school and may be unable to identify which are particularly pressing or urgent. Our intervention makes citizens aware of school performance and problems via community discussion and deliberation. Similarly, parents who are more aware of their children's education may be more compelled to take action.

2. Knowledge of and ability to contact policy actors. Even if citizens are able to identify a pressing school issue, they may not have information to act to resolve these issues in their village. Our pilot work suggests citizens lack information on who to contact and how to contact them. Our intervention thus provides citizens with information of their rights as citizens, contact information for policy actors, and guidance on how to create a message and deliver it to a policy actor.

3. Ability to take collective action. Approaching a policy actor as a group, rather than as an individual, may have greater influence on the policy actor's willingness to take action to resolve an issue. Citizens are often unable to coordinate with others in the village on education issues, and may similarly face coordination issues in approaching a policy actor in a collective manner. Our intervention helps citizens overcome the coordination issue by creating a space with a time and venue and agenda to meet and act collectively.

4. Vote choice, trust in state and self-efficacy. Following collective action, citizens who successfully reach out to a policy actor may feel empowered and self-efficacious. Positive interaction with a policy actor may increase citizen trust in state. On the other hand, if policy actors are not effective at resolving issues, citizens may be compelled to change their vote for a different party in future elections.

5. School quality. If citizens are successful in acting to solve to problems at their school, this may create a virtuous cycle that encourages them to continue to advocate for change on other school issues, increasing overall school quality.

4 Analysis

4.1 Main specification

Our basic treatment effects specification that captures the impact of our intervention is:

$$y_{vi} = \alpha + \beta T_{vi} + \theta + \epsilon_{vi} \tag{1}$$

where y_{vi} is the outcome of interest in village v at endline, measured at endline, of individual i, T_v is a treatment indicator that takes the value 1 for treatment villages, and 0 otherwise, θ theta represents randomization strata fixed effects, and ϵ_{vi} is the error term.

We will also run a specification with controls:

$$y_{vi} = \alpha + \beta T_{vi} + X_{iv}\gamma + \theta + \epsilon_{vi} \tag{2}$$

where X_{iv} is a set of control variables. We intend to separately test whether there are differences by gender, actor type and facilitation, using analogous specifications.

While our primary specification is an ITT, one can also imagine an analogous LATE specification where our treatments act as an instrument for inducing actions taken by individuals. We have two broad choices of such actions: (i) any action taken by an individual and (ii) any collective action taken by the individual. Given we are unsure how frequent either of these are in the control group and what our treatment impact will be on each, we will utilize the results of Phase I to determine which of these LATEs is relevant in our context.

We are powered up to detect combinations of any two intervention variations at any given time, including Facilitation x Gender, Gender x Actor and Actor x Facilitation. While we are not planning on running a fully interacted specification given that our sample and budget constraints imply that it is less likely that all comparisons in a fully interacted version will have sufficient statistical power, in the interest of transparency and in line with Muralidharan, Romero and Wuethrich (2022), we will also run a "long" fully interacted model:

$$y_{vi} = \alpha + \beta_1 T_{vi}^U + \beta_2 T_{vi}^P + \beta_3 T_{vi}^F + \beta_4 T_{vi}^F T_{vi}^P + \beta_5 T_{vi}^U T_{vi}^P + \beta_6 T_{vi}^F T_{vi}^U + \beta_7 T_{vi}^F T_{vi}^P T_{vi}^U + X_{iv}\gamma + \theta + \epsilon_{vi} \tag{3}$$

where T^F is a treatment indicator that takes the value 1 for villages assigned to female participation treatment and 0 for villages assigned to male participation treatment, T^P is a treatment indicator that takes the value 1 for villages assigned to politician policy actor treatment and 0 for villages assigned to bureaucrat policy actor treatment, and T^U is a treatment indicator that takes the value 1 for villages assigned to unfacilitated treatment and 0 for villages assigned to facilitated treatment.

We will also run versions of our models selecting covariates to maximize power: first, we will conduct double-lasso estimation for optimally selecting controls, identifying potential instrumental variables, and avoiding specifications search (Chernozhukov et. Al., 2016). We will thus use double-lasso to optimally select controls and improve precision

by reducing standard errors.

4.2   Treatment effect heterogeneity

We will use two different methods to examine treatment effect heterogeneity. First, we intend to examine treatment effect heterogeneity on our main outcomes along the following pre-specified dimensions:

1. How compelled citizens are to take action. This is a function of interest in their children's education and how severe they find the issue.

2. How capable citizens are to take action. This is function of interest and experience in politics.

3. The gender of the school in the village.

4. Effectiveness of policy actor (based on perceptions and measures of political competition and/or bureaucratic quality)

Secondly, we will use a "split-sample approach" using a randomly selected sub-sample and machine learning (Anderson and Macgruder, 2017) to complement this analysis. In this case, we will randomly select 20% of our sample from both Phase I and II, then use machine learning to identify relevant margins of heterogeneity. Once we have identified these dimensions in our learning sample, we will examine their effect on our main outcomes in our withheld sample.

In addition, we may also examine heterogeneity of treatment effects based on types of problems mentioned at baseline. If the problem selected to be resolved at the village level in treatment villages matches the problem an individual mentions at baseline, this may effect how much an individual cares or makes effort to resolve the problem collectively. Because village-level problem selection is endogenous (and village-level problem selection does not happen in control villages) we will construct a predicted measure of the likelihood that an individual's problem at baseline might be selected based on baseline data across villages.

At the same time, some problems selected for action during treatment might be more difficult to resolve than others. For example, building a new school is a more expensive/bureaucratically complex request than getting a water cooler at a school, and thus may influence the likelihood of problem selection and the eventual success of resolution efforts. To account for this, we will generate a difficulty index for each problem category based on its typical cost, bureaucratic complexity, and resolution timeline.

To better interpret these dynamics, we may add questions to the endline survey following analysis of midline data to help explain why respondents' problems may have been picked or not, and how satisfied respondents are in cases their problem was or was not selected at the village-level.

4.3   Spillover

Given the nature of our interventions, we also intend to study the potential spillovers our interventions may create. One potential way our interventions may generate spillovers is through geographical proximity. To explore this will examine outcomes in control villages in geographical proximity to the treated villages.