Child-led Social Accountability, Bangladesh
Pre-Analysis Plan

Akshay Dixit, Munshi Sulaiman

May 2018

1 INTRODUCTION

The goal of the analysis described here is to identify the effects of the Child-led Social Accountability project on a range of outcomes.

The plan pre-specifies the analysis that will be conducted, before comparing outcomes between treatment and comparison groups. It outlines the intervention, evaluation design, data sources, hypotheses and outcomes of interest, and the impact estimation strategy.

By committing to a pre-specified analysis plan we hope to minimize issues of data mining and specification searching. The pre-analysis plan serves the dual purpose of ensuring the endline data collection tools are sufficient for the planned analysis. This plan was submitted prior to the start of endline data collection, but after baseline data collection and the implementation of the intervention.

2 OVERVIEW OF THE INTERVENTION

The overall objective of the Child-led Social Accountability (CLSA) project is to contribute to the realization of children's rights to health and education through increased accountability and responsiveness of primary health and education service delivery in Bangladesh.

The specific objective is to design and test a child led social accountability framework, to improve the accountability and responsiveness of primary health and education improved service delivery.

The following activities were envisaged to achieve this objective:

a. Stakeholder mobilisation

This component involves introducing the project to the key "audience" (including local government officials and representatives, service providers, youth, and children) and encouraging them to participate. Essentially it lays the groundwork for the project – explaining to the audience the need for an accountability intervention centred on young people, and building a sustained rapport with those who express willingness to join the implementation.

b. Service assessments & Interface

i. Information: Children are informed of their rights and entitlements, with a focus on those specific to education and health.

ii. Social accountability tools: Children, service providers, and local government representatives co-design social accountability tools to monitor service delivery

c. Accountability for service improvement

i. Action Plans: Children discuss their issues with service providers, and children and service providers agree on certain actions that providers will take (within a given timeframe) to improve the quality of service delivery

ii. Follow-up on Action Plans: Children follow-up on the progress against the Action Plan on a bi-weekly basis

iii. Refresher information sessions for children

Clarifications regarding implementation at service delivery points

Two facets of project implementation require clarification. These are as follows:

a. **Child/adolescent monitors (Participation groups)**

One part of Stakeholder Mobilisation was identifying children who are willing to track service delivery through a participatory process. A "participation group" is simply a group of these child monitors. These groups are formed at the facility-level, so there is a distinct participation group for each of the 62 facilities working with the project. Broadly, the process involves two steps:

i. Children are given background information on the project, and encouraged to participate in the project.

ii. Those children who volunteer to participate are registered as a part of the "participation group" for that particular facility.

For education facilities, school students formed the participation group. For health facilities, children living in the vicinity of clinics/hospitals/dispensaries form the participation group.

b. "Revolving door" participation (Roster groups)

As mentioned above, for each facility, there is a group of "children/adolescent monitors". These children were randomly split into 3 groups (called "rosters"), which took turns to monitor and track service delivery. Each group participated for a maximum of 4 months. And to ensure coherence and

continuity, only after one roster group's participation ended did another group's participation begin. Dividing into three smaller groups in this manner helps ensure that each individual child has an opportunity to engage closely with service providers.

3 THEORY OF CHANGE AND RESEARCH QUESTIONS

3.1 THEORY OF CHANGE

Increased accountability and responsiveness of primary health and education service delivery, as a result of empowered, well-informed children and communities claiming their health and education entitlements through social accountability mechanisms, will contribute to the realization of children's rights to health and education.

Accountability and responsiveness of service delivery

This encapsulates two main changes:

- Increased accountability between local government/service providers and citizens (in particular children and their communities): The project aims to put forward a child-centered social accountability system that ensures constructive dialogue between children, their communities and service providers.
- Increased responsiveness around health and education service delivery: Targeted children and their communities should receive the services they are entitled to, in the specified time-frame. The quality of services should meet any existing specified quality standards/benchmarks, in addition to fulfilling the users' requirements. In case service provision is not possible at all, or quality of service provision is difficult to uphold, for any reason, this reason should be communicated clearly to the recipient.

…Under the following assumptions:

- Sustained buy-in from the local government and service providers, for the duration of the project.
- Targeted service providers are aware of the entitlements of service recipients.
- Targeted service providers are willing to engage in social accountability mechanisms and listen to feedback and demands from users (children and their communities)

- Targeted service providers are not coopted by local elites to prioritize their health and education service delivery needs
- Targeted service providers have a measure of autonomy over the financial resources and administration of their respective facilities.
- Stable political environment and security situation over the course of the project.

Empowered, well-informed children and communities, claiming their health and education entitlements

This encapsulates two main changes

- Targeted children are better informed of their rights and entitlements regarding disclosure of information on services and health and education services' entitlements
- Targeted children are empowered and capable of expressing their views/concerns and make services' related demands through constructive means. Further, children are willing to help improve the quality of service provision in their respective communities, through accountability mechanisms

…Under the following assumptions:

- Sustained buy-in from the local government, for the duration of the project.
- Children are willing to take part in social accountability mechanisms and make demands, without fear of reprisal
- Sustained children's participation/motivation over the course of the project and beyond.
- Stable political environment and security situation over the course of the project.

3.2 RESEARCH QUESTIONS

This evaluation will seek to answer the following key research questions, which form the main hypotheses that the study will test:

i. What is the effect of the intervention on access to and quality of education & health services for children?

ii. What is the effect of the intervention on children's well-being?

iii. What is the effect of the intervention on children's agency?

iv. What is the effect of the intervention on children's democratic values?

In addition to the data collected for the endline evaluation, we will also examine the above questions using some qualitative data and process monitoring data (from the intervention) to complement the findings from the analyses described below. This pre-analysis plan, however, focuses solely on the quantitative data.

4 IMPACT EVALUATION DESIGN

The design of this impact evaluation study relies on randomly assigning a set of 62 education/health facilities to one of two groups. This section describes the specifics of the randomization procedure, the treatment groups, and timing of data collection.

4.1 EXPERIMENTAL GROUPS

Following the intervention components listed in Section 2, the difference between the treatment and comparison groups can be illustrated as follows:

Treatment	**Comparison**
Stakeholder mobilisation	Stakeholder mobilisation
Service assessments and interface: i. Information ii. Social accountability tools	Service assessments and interface: i. Information
Accountability for service improvement i. Action Plans ii. Follow-up on action plans iii. Refresher trainings	

The research design does not involve a pure control group, largely for the purpose of identifying the "value-add" of the accountability tools and framework. The above set-up of the treatment and comparison groups allows for a more explicit evaluation of how the child-led social accountability

framework put in place might facilitate increased constructive dialogue between citizens and duty-bearers, and improvements in service delivery. The accountability framework is the core of the project, and the aspect that is of greatest interest to us.

4.2 SAMPLE SIZE

The intervention involved a total of 2,861 children, working with 62 facilities. By type and location, the facilities are split as follows –

Type

- Education – 36 facilities
- Health – 26 facilities

Location

- Satkhira (rural areas): 36 facilities, 18 education and 18 health
- Dhaka (urban slums): 26 facilities, 18 education and 8 health

4.3 RANDOMIZATION

Assigning facilities into one of two groups

The 62 facilities were randomly assigned to the two groups via paired random assignment. This involved two steps:

- The 62 facilities were divided into 31 pairs, with the two units in each pair matching closely on a list of baseline characteristics (see Appendix).
- In each of the 31 pairs, one of the facilities was randomly assigned to treatment, and the other to comparison.

Dividing children into roster groups

The group of participating children in the treatment group was randomly divided into 3 groups of approximately equal size[1] for each facility.

[1] This randomization was stratified by age (younger/older than the median age), gender, and whether the child's guardian completed at least primary school.

The full sample of 1451 children associated with treatment group-facilities was hence divided into three groups of sizes 485, 482 and 484 respectively, which took turns to participate in the intervention. Children associated with comparison group-facilities were not divided into groups.

4.4 SURVEY ROUNDS & TIMELINE

As mentioned above, 2861 children participated in the intervention. We plan to contact all of these children to conduct the endline survey.

We will also be conducting an endline survey with representatives of the 62 education/health facilities who participated in the intervention.

In addition, to understand the general impact on the intervention areas, we will collect data on a set of randomly-sampled children who did not participate in the intervention. For education facilities, we will record certain administrative data (e.g. test scores from the last standardized examination that they took) on 30 children per sample facility, randomly sampled from among children who are enrolled at the facility. For health facilities, we will record Patient Reported Experience Measures (PREMs) from caregivers/guardians of 10 children per sample facility. These will be randomly sampled from among children who have visited the sample health facility in the last month for healthcare.

Survey round	**...Focused on research questions (RQs)**	**Sample size**	**Timing**
Baseline survey (I)	RQ 1	3,141 children[2] 62 service providers	Nov 2016
Baseline survey (II[3])	RQ 2 and 3	2,861 children	April 2017
Intervention Period: June 2017 to May 2018			

[2] Of these, 2,861 children participated in the intervention.

[3] The baseline survey was broken up into two rounds to break up one long survey into two, which helped alleviate problems such as survey fatigue.

Mid-line survey	RQ 1 and 3	892 children[4] 62 service providers	Dec 2017
End-line survey (I)	RQ 1, 2, 3 and 4	2,861 children participating in the intervention 62 service providers PREMs on 260 children not participating in the intervention	July 2018
End-line survey (II)	RQ 3	Exam scores/attendance of 1080 children not participating in the intervention	November 2018[5]

4.5 BASELINE EQUIVALENCE

We verified that the treatment and comparison groups do indeed look similar on a host of baseline characteristics. This is true for both facility characteristics, as well as characteristics of the participating children associated with the facilities. Overall, the differences between the groups tend to be very small in magnitude and statistically not significant.

Similarly, we also verified that the three roster groups of children associated with treatment group facilities look similar on baseline characteristics. Overall, the differences between the groups tend to be very small in magnitude and statistically not significant.

4.6 PANEL DATA

The endline survey is a follow-up to the baseline survey. While the survey instruments we intend to use at endline are in some ways different from the ones used at baseline, we anticipate having two separate panel data-sets (from children and from service providers) for most (though not all) of our outcomes.

---

[4] The midline survey was conducted with the 967 children at baseline who were assigned into rosters one and three. Successfully follow-up interviews were conducted with 892 of these children.

[5] This is in accordance with the results timeline for the relevant examinations in Bangladesh

5 IMPACT ESTIMATION STRATEGY

5.1 REGRESSION MODEL

The treatment was randomized at the facility level, but our outcomes of interest are measured both at individual level (e.g. agency, well-being, perceptions of quality of service delivery by the children) and at facility level (e.g. quality of service). Our main specification for measuring impacts on child level outcomes (e.g. their agency) will be

$$y_{ijendline} = \alpha + \beta_1 T_j + \beta_2 PAIR_j + \delta_1 y_{ijbaseline} + \varepsilon$$

Where $y_{ijendline}$ is the outcome of child i linked to facility j at endline, and T_j is the dummy for treatment assignment of facility j, In this specification, δ_1 controls for the baseline value of the corresponding outcome indicators (if measured at baseline) and β_1 is our impact estimate. In this this base specification, ε is the error term that is clustered at facility level, and the variable $PAIR_j$ is a vector of dummy variables that indicate the randomization pair of facility *j.* In addition to this specification, we will analyze the data using the following addition of dummies for roster assignment to measure if early participation makes any difference in children's agency and wellbeing, where $R1_{ij}$ and $R2_{ij}$ are dummies that indicate if the child was assigned to roster 1 and roster 2 respectively (with roster 3 as the base category).

$$y_{ijendline} = \alpha + \beta_1 T_j + \beta_2 PAIR_j + \beta_3 R1_{ij} + \beta_4 R2_{ij} + \delta_1 y_{ijbaseline} + \varepsilon$$

We will use randomized inference method of calculating exact p-values for the above specifications. In case of attrition being too high (e.g. above 10%) or differential attrition between the treatment and comparison groups, we will use inverse probability weights of tracking for adjusting our estimates.

For the outcomes measured at facility level, we shall use the following specification

$$Y_{jendline} = \alpha + \beta_1 T_j + \beta_2 PAIR_j + \delta_1 Y_{jbaseline} + \varepsilon$$

Where the endline values of the outcomes will be compared between treatment and comparison facilities by controlling for their baseline estimates (if the corresponding outcomes were measured

at baseline). Similar to individual level outcomes, we shall calculate exact p-values using randomized inference approach.

5.2 ADJUSTMENT OF P-VALUES FOR MULTIPLE INFERENCE

We are measuring impact separately on three distinct "families" of outcomes –

i. Access to and quality of service delivery, and children's well-being – for children who participated in the intervention (RQ 1 and 2)
ii. Children's agency and democratic values (RQ 3 and 4)
iii. The general impact on outcomes of children in intervention areas (RQ 2).

The first set of outcomes pertains to education and health service delivery, and well-being of children who participated in the intervention. The second set of outcomes pertains to the impact on the participating children's democratic values and empowerment. The third set of outcomes draws on a round of data collection separate from the first two, with a different sample.

Each of these families has multiple outcomes, and there is a chance of over-rejection of the null hypothesis of no impact owing to multiple hypothesis testing. Checks against multiple hypothesis testing include ex-post adjustment of p-values, and reducing the number of hypotheses tested ex-ante.

The approach we will take for the outcomes within each family is to control the False Discovery Rate (FDR), which limits the expected proportion of rejections within a hypothesis that are Type I errors (Benjamini, Krieger, and Yekutieli 2006[6]; Anderson 2008[7]; Casey, Glennerster, and Miguel 2012[8]). This will guard against false rejections of the null hypothesis for key outcomes and therefore against falsely declaring statistically significant the overall impacts of the project. We will not adjust p-values for any exploratory analyses that we do, since those results will be used to

[6] Yoav Benjamini, Abba M. Krieger, and Daniel Yekutieli, "Adaptive linear step-up procedures that control the false discovery rate", *Biometrika*, 93, 3 (2006): *pp.* 491-507

[7] Michael L. Anderson, "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, no. 484 (December 1, 2008): 1481–95.

[8] Katherine Casey, Rachel Glennerster, and Edward Miguel, "Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan," *The Quarterly Journal of Economics* 127, no. 4 (November 1, 2012): 1755–1812.

improve our understanding of the project's impact[9], and not to judge the success of the intervention.

Adjusting p-values for multiple hypothesis testing reduces the power to detect effects for each outcome individually, and hence it is important to limit the number of outcomes considered. To that end, certain outcomes will be grouped into indices (see Section 6.1 for the list of outcomes and indices). Index construction is described in Section 8.

6 OUTCOMES OF INTEREST

This section describes the various outcomes that will be used in the quantitative analysis of the project.

Here, we specify the set of outcomes that will serve as the main measures of overall project impact in this endline analysis. These specifications were chosen to address all of the key research questions, minimize the likelihood of false claims of impact, and maintain the ability to detect any project impacts, if they occurred. The results from these analyses will be used to make judgments on overall project impact.

We do recognize that new hypotheses may emerge during the course of data collection and analysis, especially if we see unanticipated realizations of the data. We intend to explore such hypotheses. We leave open the possibility of applying different tools or identification techniques to test any emerging hypotheses, as warranted by the realization of the data that we see. We will, however, be explicit about which analyses were pre-specified and which ones emerged later.

Note that while many outcomes apply to both education and health facilities (e.g. outcomes on empowerment and democratic agency, the presence of feedback mechanisms, general quality of facility infrastructure, etc.), some well-being outcomes are specific to education or health. For such outcomes, analysis will be performed on the relevant sub-set of the data. This distinction has been clarified below, where necessary (see Section 6.2).

[9] This is discussed further in Section 6.

6.1 RESEARCH QUESTION 1: ACCESS TO & QUALITY OF SERVICE DELIVERY

Sr. No.	Outcome	Definition
1.	Access to services for deprived groups of children	Index of access associated with the following categories, as reported by the sample service providers: • Percentage-bracket of working children using the facility's services • Percentage-bracket of children with disability using the facility's services • Percentage-bracket of children living in slums (for Dhaka) OR remote rural areas (for Satkhira) • Percentage-bracket of children from poor families (defined as having a family income of less than roughly BDT 5000 per month) The "percentage-bracket" referred to above means that the responses are recorded on a scale of 1 to 7, as follows: 1. No children in this category 2. Less than 1 % 3. Between 1 and 2% 4. Between 2 and 5% 5. Between 6 and 10% 6. Between 11 and 15% 7. More than 15%
2.	Feedback mechanism index	This consists of two components: • Whether the respondent reports having a mechanism to provide

		anonymous feedback on any teacher/health service provider • Whether the respondent reports the feedback mechanism as functional
3.	Infrastructure Index	This index will be constructed separately for Education and Health facilities. For Education facilities, this includes the following components: • Respondent's overall rating of the school infrastructure, on a scale of 1 (very good) to 5 (very bad). • Respondent's incremental rating of the school infrastructure, on a scale of 1 (improved a lot) to 5 (worsened a lot). • Whether the respondent reports the facility as having: o Enough toilets for all students o Handwashing facility o Safe drinking water For Health facilities, this includes the following components: • Respondent's overall rating of the health facility infrastructure, on a scale of 1 (very good) to 5 (very bad). • Respondent's incremental rating of the health facility infrastructure, on a scale of 1 (improved a lot) to 5 (worsened a lot). • Respondent's rating of the facility equipment, on a scale of 1 (very good) to 5 (very bad) • Whether the respondent reports electricity problems at the facility. • Whether the respondent reports the availability of safe drinking water at the facility
4.	Provider engagement & effort index	This index will be constructed separately for Education and Health facilities.

		For Education facilities, this includes the following components: • Number of teaching staff vacancies, measured relative to the total number of teaching staff at the facility • Number of non-teaching staff vacancies, relative to the total number of non-teaching staff at the facility • Total number of students currently enrolled at the facility, relative to the total number of teaching staff at the facility. • Number of days any of the respondent's teachers were absent in the last month when the facility was open For Health facilities, this includes the following components: • Number of healthcare provider vacancies, measured relative to the total number of healthcare staff at the facility • Number of non-healthcare staff vacancies, relative to the total number of non-healthcare staff at the facility • Number of times the respondent or his/her family were not seen by a healthcare provider upon visiting the facility • Respondent's rating of the health facility staff behavior, on a scale of 1 (polite and respectful) to 4 (abusive) • Usual waiting time (in minutes) to see a healthcare provider, as reported by the respondent
5.	Children's perceptions	This index will be constructed separately for Education and Health facilities. For education facilities, this includes the respondent's perceptions of 13 aspects of the quality of services at the facility, each measured on a scale of 1 (strongly agree) to 5 (strongly disagree).

		For health facilities, this includes the respondent's perceptions of 14 aspects of the quality of services at the facility, each measured on a scale of 1 (strongly agree) to 5 (strongly disagree).

6.2 RESEARCH QUESTION 2: CHILDREN'S WELL-BEING

As mentioned above, well-being outcomes are split into two groups: one pertaining to children who participated in the intervention, and one aimed at measuring general impact on children in the intervention areas, measured for a random sample of children who did not participate in the intervention. In that order, there are two separate tables below.

Sr. No.	Outcome	Definition
1.	Experience of harassment/aggression	Whether the respondent experienced any physical punishment, or psychological aggression, or sexual harassment in the past two months
2.	Reporting of harassment/aggression	Whether the respondent shared his/her experience of physical punishment, or psychological aggression, or sexual harassment with anyone
3.	Corporal punishment (specific to Education facilities)	The frequency with which teachers at the respondent's school physically punish students for any reason, on a scale of 1 (Frequently) to 4 (Never). This is specific to education facilities.
4.	Student attendance (specific to Education facilities)	Number of days the respondent did not go to school in the last month when the facility was open. This is specific to education facilities.
5.	Child health (specific to Health facilities)	Number of times the respondent fell sick/injured in the last 6 months.
6.	Utilization of health services (specific to Health facilities)	Number of times the respondent visited this facility in the last 6 months, relative to the number of times the respondent fell sick/injured.

Sr. No.	Outcome	Definition
1.	Student attendance (specific to Education facilities)	Proportion of days a student did not attend school in the recently-concluded academic year
2.	Student exam scores (specific to Education facilities)	Standardized aggregate of student's performance across all subjects in the relevant National examination. Depending on the age of the child, the test could be Primary School Certificate (PSC - after completion of 5 years of primary school), Junior School Certificate (JSC - after completion of 8 years of schooling) or Secondary School Certificate (SSC - after completion of 10 years of schooling) Examination.
3.	Patient Reported Experience Measures – overall satisfaction (specific to Health facilities)	Respondent's overall satisfaction with the healthcare and its outcome. This will be constructed as an index including three components: • Respondent's overall rating of the quality of healthcare, on a scale of 1 (Excellent) to 5 (Terrible) • Respondent's willingness to recommend the facility to family/friends, on a scale of 1 (Definitely Yes) to 4 (Definitely No) • Respondent's satisfaction with the outcome of healthcare, on a scale of 1 (Completely Satisfied) to 5 (Completely Dissatisfied)
4.	Patient Reported Experience Measures – perceptions regarding various specific aspects of the quality of care (specific to Health facilities)	The aggregate of the respondent's perceptions of 7 aspects of the quality of services at the facility, each measured on a scale of 1 (strongly agree) to 5 (strongly disagree).
5.	Utilization of health services (specific to Health facilities)	The average number of patients visiting the facility per day during the past month.

6.3 RESEARCH QUESTION 3: CHILDREN'S AGENCY

Sr. No.	Outcome	Definition
1.	Grit	Respondent's score on the 8-item grit scale[10]
2.	Locus of Control	Respondent's score on a 17-item locus of control scale[11]

6.4 RESEARCH QUESTION 4: DEMOCRATIC VALUES

Sr. No.	Outcome	Definition
1.	Political Communication	Number of people (up to a maximum of 3) that the respondent talked with about political matters in the last month.
2.	Political Awareness – I	Number of questions from among the following that the respondent gives the correct answer to: • What is the name of your Union Parishad chairman/Ward Councilor? • Who is the current President of your country? • How many Members of Parliament are there in your country?
3.	Political Awareness – II	Whether the respondent reads or watches political news.

[10] Duckworth, A. L., and Quinn, P. D. (2009) "Development and validation of the Short Grit Scale (GritS)", *Journal of Personality Assessment*, vol. 91(2):166-74.

[11] 17-item scale adapted from – Nowicki, S., & Strickland, B. R. (1973) "A locus of control scale for children", *Journal of Consulting and Clinical Psychology*, vol. 40(1):148.

4.	Democratic agency – National	The perception of the respondent about how much the National government would listen to the respondent in solving a problem specific to the respondent's area, measured on a scale of 1 (a lot) to 4 (none at all).
5.	Democratic agency – Local	The perception of the respondent about how much the Local government would listen to the respondent in solving a problem specific to the respondent's area, measured on a scale of 1 (a lot) to 4 (none at all).
6.	Attitudes towards diversity	Number of statements from among the following to which the respondent responds in a manner indicative of a favorable attitude towards diversity. • I prefer to have friends who are very similar to me in terms of gender, religion, and economic status. (Yes/No) • I enjoy meeting people who come from backgrounds (i.e. gender, religion and economic status) very different from my own. (Yes/No) • Imagine two groups of children. In the first group, the children mostly come from the same background – i.e. the same religion, the same gender, and similar economic status. In this way, they are similar to each other. The second group includes children from different kinds of backgrounds – a mix of hindus and muslims, boys and girls, coming from different economic status. Which group would you want to be a part of? (Group 1/Group 2)

7 SUB-GROUP ANALYSIS

We plan to estimate the impacts of the program on a number of key sub-groups:

- The intervention areas include both urban and rural communities. We will examine whether the project impact differed between urban and rural areas.
- For service delivery outcomes common to both education and health facilities, we will examine whether the project impact differed between education and health facilities.
- For child-level outcomes, we will examine whether the project impact differed by gender (male/female), and age (younger/older than the median age).
- The efficacy of the intervention, in terms of the number/type of actions that were undertaken to improve the quality of services delivered, varies across sample facilities. We will divide the sample facilities into three groups (of low/medium/high efficacy) based on an analysis of the Action Plans undertaken at each of the facilities, and examine how the intervention's impact varied between the groups.

To conduct sub-group analyses, we will use the regression strategy described in section 5 but adding dummy variables for the sub-groups and an interaction between the treatment dummy variable and the relevant sub-group. The coefficient on the interaction will represent the difference in the impact of the program for that sub-group relative to the omitted sub-group.

8 VARIABLE CONSTRUCTION

As mentioned above, certain outcomes will be grouped into indices. Section 8.1 describes how these indices will be constructed. The remainder of this section is relevant for all variables.

8.1 INDICES

We will construct mean effect indices following the procedure outlined in Casey,

Glennerster, and Miguel (2012)[12] which follows on Kling, Liebman, and Katz (2007)[13]. The steps involved in estimating the mean treatment effect are as follows:

1. Each outcome is first oriented so that higher values represent "better" values.
2. Then, each outcome is standardized by subtracting the mean of the outcome and dividing by the standard deviation of the comparison group.
3. Missing values are imputed at the treatment assignment group mean.
4. Finally, a summary index is compiled that gives equal weight to each individual outcome component.

The aforementioned approach weights each outcome component of the index equally. Anderson (2008)[14] weights each outcome component by the inverse of the appropriate element of the variance-covariance matrix (as measured in the comparison group), which "down-weights" outcome components that are highly correlated with each other. We will check robustness using the weighted version and note any differences.

8.2 DON'T KNOW AND REFUSED

"Refused" will be coded as missing.

The treatment of "Don't know" is outcome-specific. In general it will be coded as missing, but there are two key exceptions. In the context of index formation, "don't know" will be treated as a missing value, and imputed at the treatment assignment group mean. "Don't know" in response to knowledge questions will be coded as an incorrect answer. For all other variables, if more than 30% of the values are "don't know", the variable will be dropped from the analysis.

---

[12] Katherine Casey, Rachel Glennerster, and Edward Miguel, "Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan," *The Quarterly Journal of Economics* 127, no. 4 (November 1, 2012): 1755–1812.

[13] Jeffrey R Kling, Jeffrey B Liebman, and Lawrence F Katz, "Experimental Analysis of Neighborhood Effects," *Econometrica* 75, no. 1 (January 1, 2007): 83–119.

[14] Michael L. Anderson, "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, no. 484 (December 1, 2008): 1481–95.

This will be followed generally, but we will keep the possibility open for imputing data in situations where the fraction of "don't know" responses is substantial, but not high enough to discard the variable entirely.

8.3 Missing data from Item Non-response

After recoding don't know and refused values, we will check for balance on missing values and test the sensitivity of our results to different assumptions on the missing data (due to non-response/DK, etc.). If necessary, due to various assumptions about missing data, we will create upper and lower bounds by recoding missing values for treatment as 0 and comparison as 1 and vice versa.

8.4 Outliers

Most of our outcomes are not prone to outliers – they are either scales, or indices composed of variables with pre-defined choice menus.

For the continuous outcomes, we will first check that the reason for the outliers is not data collection/entry error. If it is not, the analysis will be performed both including and excluding outliers, to check how sensitive the results are to the presence of outliers.

9 APPENDIX – CRITERIA FOR PAIR MATCHING OF FACILITIES

List of baseline characteristics used to do the pair-matching: Education facilities

Children's questionnaire

Question
How long does it take you to get to school/madrassa from home every day?
Do you have any classmates who work to earn money?
Do any of your classmates have a disability?
Do any of your classmates live in an area that is remote, or very difficult to reach from here?
Normally, how many hours a day is your school/madrassa open?
Usually, how clean are your school/madrassa toilets? (Rate on a scale of 1 to 5)
Is the water at school/madrassa safe for you to drink?
Is there a School/madrassa Management Committee for your school/madrassa?
Does your school/madrassa have a student council/student representative group?
If you have any feedback on any of your teachers, is there a system in place at school/madrassa for you to report the feedback?
Do you think students should have a say in how the school/madrassa money is spent?
Do teachers at your school/madrassa physically punish students for any reason?
School/madrassa premises are clean and hygienic. (to what extent to do you agree?)
It is necessary to take external tuitions, to do well in exams. (to what extent to do you agree?)
It is OK for teachers to sometimes physically punish students, to get them to behave properly. (to what extent to do you agree?)

Service provider questionnaire

Question
Which union/ward are you conducting this interview in?
What type of school/madrassa is this?
Is it a primary or secondary school/madrassa?
What is the total number of classrooms at the school/madrassa?
In total, how many boys are currently enrolled in the school/madrassa?

In total, how many girls are currently enrolled in the school/madrassa?
Does your school/madrassa have double shift classes?
Does the school/madrassa have a library?

List of baseline characteristics used to do the pair-matching: Health facilities

Children's questionnaire

Question
Have you visited _____ facility in the last 6 months?
Do patients have to pay a fee for the doctor when they visit this health facility?
Have you ever seen/heard of children from remote, very hard-to-reach areas visiting this facility?
Have you heard people – your family, neighbors, friends, teachers, etc. – complain about how long they have to wait before seeing a doctor/nurse?
Does it sometimes happen that the health facility is closed, when you or anyone you know (family, friends, neighbors, etc.) go to the health facility for a consultation?
If you have any feedback on the health services/staff, is there a system in place to report this feedback?
Have you ever received any information/services from the health workers of this facility?
Do you think this health facility has enough money to provide good quality healthcare?
How long does it take you to travel from home to this health facility?
Does it sometimes happen that you or your family are not seen by a doctor/nurse when you go to the health facility?
Are services for children (e.g. vaccinations, growth monitoring, etc.) available at this facility?
How clean are the facility toilets? (Rate on a scale of 1 to 5)
To get good quality healthcare, one needs to have good relations with doctors/health officials. (to what extent to do you agree?)
When the clinic/hospital is making a decision about the services they provide, community members are consulted about it. (to what extent to do you agree?)
Health staff pay attention to a child's problems only when he/she is with parents or adults. (to what extent to do you agree?)
Wealthy people get better quality treatment at this facility. (to what extent to do you agree?)

Service provider questionnaire

Question
Which union/ward are you conducting this interview in?
What type of healthcare facility is this?
Does the facility have its own building to run its activities?
Is it registered with the government?
During the past month, what was the average daily number of patients visiting the facility?
How many patients visited the facility yesterday?
How many male doctors currently work here?
How many female doctors currently work here?
Are there any vacant positions?
Does the facility have basic essential equipment for children?
How do you rate the availability of essential medicine for children? (on a scale of 1 to 5)