Secondary School Readiness Program (SSRP): A Pre-Analysis Plan

Sabrin Beg, University of Delaware
Anne Fitzpatrick, University of Massachusetts Boston
Jason Kerwin, University of Minnesota
Adrienne Lucas, University of Delaware
Khandker Wahedur Rahman, University of Minnesota

December 20, 2019

1. Introduction

This pre-analysis document describes our plan for analyzing a randomization evaluation of the Secondary School Readiness Program (SSRP). The version of the SSRP, or "Utkarsh" program that we evaluate is the implementation in Odisha state, India. This document details our primary outcomes of interest, the regression specifications we plan to use to measure program effects, and our plan for multiple hypothesis testing adjustments.

In addition to the treatment effect on the primary outcomes of interest, we will explore specific channels or mechanisms through which impact was achieved, or not.

This study is conducted with our partners, People for Action (PFA) and the Odisha Department of School & Mass Education (SME). People for Action, an NGO that created the original SSRP program, designed all program materials. SME implemented all aspects of the program with technical assistance by PFA. All data collection was conducted by Jameel Poverty Action Lab South Asia (JPAL-SA).

2. Intervention

The "Utkarsh" or SSRP intervention is a government-led remedial and grade-level program for 9th grade students[1], and focuses on 4 subjects: Odia, English, Math and Science. The program consists of three phases, each targeting a different "learning level", designed to bring the lowest level learners to grade level and improve the grade-level knowledge of all students within a single school year. The first phase, the Foundation Camp (FC), targets the weakest students, those with a 3rd grade or lower learning level, and builds their foundational skills. The second phase, the Supported Learning Phase (SLP), adds medium-level students, or those with learning level at approximately 5th grade, and develops applications of foundational knowledge. The final phase, Consolidation Camp (CC), includes the entire class and focuses on grade-level material in preparation for the transition to 10th grade and the Class 10 board exams. These sessions operate within the typical school day with existing teachers.

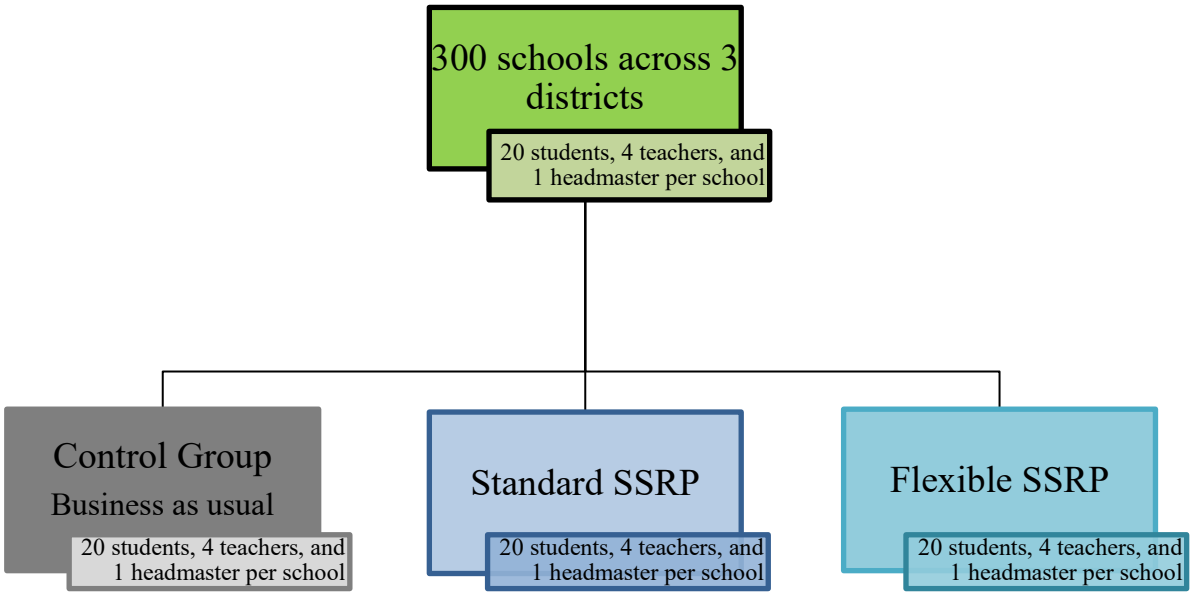
We evaluate two models of the SSRP intervention, the *Standard SSRP* and the *Flexible SSRP*. The Standard SSRP includes a fixed curriculum for each of learning phases. Teachers are provided materials for the SSRP lessons during the training and implement the program's topics and lesson plans according to the pre-specified schedule. The Flexible SSRP differs from the standard version by offering more autonomy to teachers to adapt topics and/or timelines according to the needs of their students. Specifically, in the second and longest phase of the program, the Supported Learning Phase (SLP), teachers can choose which SSRP lessons to implement and in which order. This is facilitated through the use of a worksheet which tracks the lessons that teachers choose to implement. The other two phases of the program (FC and CC) are implemented as usual.

[1] 9th grade is referred to as Class 9 in the local school system, and is followed by Class 10 (10th grade). Students take a set of high-stakes exams at the end of 10th grade, the Class 10 Board Exams, which determines their onward educational enrollment.

3. Research Design and Data

3.1 Experimental Design

This randomized controlled trial is in 2 districts in Odisha (Jajpur and Dhenkenal). This study will operate in 300 schools within these districts. The SSRP evaluation has three study arms (two treatment groups and one control group). The study design is pictured in Figure 1.



At each school, we randomly selected up to 20 students to participate in study. At each school, we interview 4 teachers and the headmaster or in-charge of the school. Due to low enrollment and vacancies, the actual number of respondents per school may differ.

Treatment 1, (T1) Standard SSRP — (100 schools)

Treatment 2 (T2), Flexible SSRP — (100 schools)

Control (C) —9th grade teachers and headmasters continue as usual, receiving no additional training or resources. (100 schools)

Impact of the Intervention: Our design allows us to test three hypotheses:

H01. Does Standard SSRP improve primary and secondary outcomes of interest? (T1 vs. C)
H02. Does Flexible SSRP improve primary and secondary outcomes of interest? (T2 vs. C)
H03. Is the effect of Standard SSRP significantly different from Flexible SSRP? (T1 vs. T2)

3.2 Power Calculations

Our primary outcome is student performance on the four program subjects. Assuming a conservative estimate of 30 percent attrition and a 0.4 increase in R-squared with the inclusion of baseline test scores, we estimate that our minimum detectable effect (MDE) is 0.17 standard deviations for Science and Odia and 0.18 standard deviations for Math and English. This is based on controlling for all available baseline scores when analyzing effects on endline test scores. In sum, we are adequately powered to detect meaningful differences in test scores between study arms – both between each of the treatment arms and the control arm, as well as comparing the treatment arms with one another.

3.3 Sample Selection and Data

School Selection:

We randomly selected 300 schools from the study districts, excluding schools that did not use the official state language as the instructional language, were governed by the SC/ST Development Department, did not have any enrolled 9th grade students or were schools for special needs students. Among selected schools, we then implemented a stratified randomization to result in three comparable groups. We stratified the random assignments to study arm by district, 9th grade enrollment, teachers per student, Class 10 board exam passing rate in the previous year, and distance from the district headquarters.

Data Collection

We conduct three waves of data collection through surveys and tests in our study schools: baseline, midline/monitoring, and endline. In this document a "wave" refers to one of those three visits to a school. The timeline of data collection is as follows:

- July 2019: Baseline survey conducted
- August 2019: Teacher training and headmaster orientation
- September 2019: FC phase monitoring visits (100 schools)
- September-October: Phase-II teacher training
- October-November 2019: SLP Phase monitoring visits (200 schools)
- December 2019-February 2020: Endline survey conducted.

The endline surveys will take place from December 2019 to February 2020, but school holidays will take place between late December and early January. Thus, we randomly selected 60 schools to do the endline prior to the holidays, with the remaining 240 taking place afterward.[2]

To mitigate attrition at the endline, we plan to conduct two rounds of mop-up visits to schools to find students are teachers who were missed during our initial attempt. The first mop-up visit will happen at all 300 schools. The second mop-up visit will happen at 150 schools, selected at random.

[2] For the early endline visit, we randomly picked 9 strata from Dhenkanal district and selected all the schools in each strata into the early visit group. The early endline schools have an equal number of treatment and control schools.

To improve balance for this random assignment, we conducted a re-randomization exercise. We created 1200 potential random assignments. For each assignment, we ran individual t-tests on a set of baseline variables for students, teachers, and headmasters/schools (using the headmaster survey).[3] For each baseline variable we conducted six t-tests of the equality of the means, examining each pairwise comparison of study arms for both the mop-up group and the non-mop-up group. This yielded a set of 204 t-statistics, which we then took the absolute value of. We then picked out the largest (i.e. least-balanced) of these. Our selected randomization uses the random assignment where the selected t-statistic was the lowest, i.e. the one where the least-balanced balance test looked the most-balanced.

Study Participants

This study has three respondent types: headmasters, teachers, and students. Each respondent type will complete a different survey, and students also took a set of standardized examinations. Enumerators conducted all surveys in Odia and entered responses directly into tablets. Participation by individuals was determined by their school's inclusion in the study, and (in the case of students) random selection. Below we describe the different kinds of participants, the expected and actual sample size at baseline, details of how they are included in the sample and a brief overview of the information collected:

1. **Students:** We selected the study students for baseline from those currently enrolled in 9th grade and present during the day of the baseline survey visit. We first randomly selected a section (if a multi-section school) and from that section randomly selected 20 students to participate in the study. If a section or school had fewer than 20 students, all students were included. Our baseline sample consists of 5,745 of a possible 6000 Students.[4] We administered a short one-on-one demographic survey, and a group assessment on their Odia, Math, and English skills.[5]

   At the midline/monitoring visits we collected data on attendance of the study students. We will attempt to re-interview and test all the baseline students at the follow-up stage.

[3] The student balance variables we used were: percent scored at baseline in English, Math, and Odia tests, percent non-response at baseline in English, Math, and Odia tests, a dummy variable for being male, age, number of siblings, a dummy variable for being Hindu, dummy variables for belonging to general caste, scheduled castes/scheduled tribe, other backward castes, other castes, and a dummy variable for speaking Odia as the primary language. The teacher balance variables we used were: age, a dummy variable for being male, a dummy variable for having a bachelor's degree of above, number of years as a teacher, number of years as a teacher in the surveyed school, a dummy variable indicating whether works at another schools, days absent from work, days late to work, number of 9th grade sections taught by the teacher, and a dummy variable for having an outside job. The headmaster and school balance variables we used were: age of the headmaster, a dummy variable for the headmaster being male, a dummy variable indicating whether the person in-charge of the school is designated as headmaster, a dummy variable for the headmaster begin in-charge every day, a dummy variable for the headmaster having a bachelor's degree or above, number of years spent as a teacher, number of other schools where this person is the headmaster or in-charge, number of posts for teachers in the school, number of posts for teachers in the school that are filled, number of boys enrolled in the 9th grade, number of girls enrolled in the 9th grade, a dummy variable for the school participating in Utthan, a remedial program for lower grades that was also available to school in the study districts.

[4] Note that some students left midway through the school day during the baseline enumeration and so there are unequal numbers of students across the surveys and the 3 subject exams.

[5] Due to limited time during the school day, we did not administer the Science exam during the baseline assessment.

2. **Headmasters:** In-person surveys were conducted with the headmaster or in-charge of each of the study schools. The headmaster survey covered a range of topics from information on their school and characteristics, to their personal background and school management practices. Our baseline sample includes 299 of a possible 300 Headmasters.

3. **Teachers:** Surveys were conducted in-person and covered a range of topics from information on their personal background to their workplace experience and challenges. At baseline all 9th grade teachers who taught Odia, English, Math, or EVS in the 300 study schools were targeted for study inclusion. Our baseline sample consists of 831 of a possible 1200 teachers due to teacher vacancies, teacher absenteeism, and refusals. The teachers surveyed at midline and endline are not always the same as those surveyed at baseline because of turnover and because some teachers were missed at baseline. We will test for selective attrition from and entry into the teacher sample in the midline and endline waves.

4. Outcomes and Econometric Specifications

In this section we first provide our primary econometric specification and then list the specific outcomes of interest.

4.1 Regression Specification

As this is a randomized controlled trial, the estimating equation is straightforward:

$$Y_{isj} = \beta_0 + \beta_1 StandardSSRP_s + \beta_2 FlexibleSSRP_s + \delta' X_{is} + \gamma_j + d_{isj} + w_{isj} + \epsilon_{isj} \quad (1)$$

which is an ANCOVA specification where Y_{isj} is the outcome of interest measured at endline for respondent i in school s in stratification cell j. *StandardSSRP* and *FlexibleSSRP* are dummy variables indicating the randomly assigned treatment status of the school. These indicators are mutually exclusive. X is vector of control characteristics, including the baseline value of the outcome variable if measured, and the wave of survey (if the outcome is measured at multiple waves).[6] Other controls included in X may vary depending on the outcome Y and respondent i, and are specified for all of the outcomes listed below. γ is a fixed effect for our stratification cells as listed above, d is a day-of-week fixed effect, and w is a week fixed effect. For outcomes measured during monitoring visits, we will use a pooled sample across all three phases of visits (FC, early SLP, and late SLP). All standard errors will be clustered at the school level.

Tests of H1-H2 correspond to whether β_1 and β_2 are statistically different from 0. Tests of H3 corresponds to whether β_1 and β_2 are statistically different from each other. We will use an F-test to test for the equality of the estimates of β_1 and β_2.

[6] In all our specifications, if a control variable is missing for any reason, we dummy out that missing value by setting the missing values to zero and including as an additional control an indicator for the variable being missing.

4.2 Outcomes of Interest

Outcome indices

Several of the outcomes listed below are indices of multiple different underlying variables. Unless otherwise specified, all indices will be constructed using principal components analysis as follows: we will take the first principal component of the variables for the entire sample, just for the wave in question, and use the weights for the first principal component.

If one of the underlying variables is missing, we omit it from both the numerator and the denominator. That is, consider an index composed of two dummy variables, one with weight 0.4 and one with weight 0.6. If both dummy variables are present, , people with missing data from one of the variables would get an index that is based on just the one available variable, with a denominator of one instead of two. We will treat "Don't Know" and "Refused" as missing.

After constructing the indices, we will standardize them by first subtracting off the mean for the entire sample and then dividing by the standard deviation for the entire sample.

Note that the student test score indices are constructed using IRT, *not* using PCA.

4.2.1 Primary Outcomes

4.2.1.1 Student outcomes:

1. Standardized test score on the English exam at endline.
2. Standardized test score on the Odia exam at endline.
3. Standardized test score on the Math exam at endline.
4. Standardized test score on the Science exam at endline.

Two notes on the test score outcomes:

a. Item response theory will be used convert each students' score to reflect their latent ability. Specifically, following the procedure used in our pilot, we construct IRT-based test scores separately for each subject and for each wave of the study, pooling data across all three study arms. We will run one-parameter logit models in Stata by using the "irt 1pl" command, and then predict the latent values of the score variable by using "predict SCORE, latent". IRT scores will then be standardized by subtracting off the control-group mean and dividing by the control-group standard deviation. We will drop any assessment questions from follow-up for which there is a negative discrimination parameter.

b. All our test score regressions, including the heterogeneous treatment effects analyses listed below, will control for student-level baseline scores of all three tests conducted at baseline.

5. Attendance during unannounced monitoring visits
    a. Attendance is a binary variable of being present or not on the day of monitoring visit.

All student regressions will also control for a dummy for being female and age.

4.2.1.2 Teacher outcomes:

The majority of the primary outcomes for teachers come from the teacher survey conducted during the midline data collection wave. We draw on two variables from the midline classroom observation and one from the endline teacher survey. The outcome variables for teachers are:

1. SSRP knowledge index, based on the following variables:
    a. Correct answer to which students are supposed to attend FC (inception and class 3)
    b. Correct answer to which students are supposed to attend SLP (inception, class 3, and class 5)

2. SSRP implementation index, based on the following variables:
    a. Indicator for conducting levelling assessment
    b. Indicator for teacher saying they are implementing correct phase of the program at monitoring visit
    c. Indicator for delivering FC to the correct set of students (inception and class 3). Set to missing if FC not yet scheduled to occur. Half-credit if teacher lists correct students but also other groups of students.
    d. Indicator for delivering SLP to the correct set of students (inception, class 3, and class 5). Set to missing if SLP not yet scheduled to occur. Half-credit if teacher lists correct students but also other groups of students.
    e. Indicator for delivering CC to correct set of students (all students). Set to missing if CC not yet scheduled to occur. Half-credit if teacher lists correct students but also other groups of students. *(this outcome is measured on the endline teacher survey)*
    f. Fraction of past 6 days that Utkarsh was implemented
    g. Indicator for teaching Utkarsh lessons on more than half of past 6 days. We replace this indicator with 1 if teacher taught Utkarsh lessons on fewer than 4 out of last 6 days but said this was because another teacher taught those lessons.
    h. Indicator for students currently using handbooks in class. *(This outcome measured on classroom observation tool)*. This variable is only defined for the teacher(s) whose lesson was observed, and is missing otherwise.
    i. Indicator for classroom having a word wall – a section of the classroom devoted strictly to vocabulary to be used during the course or chapter. *(This outcome measured on classroom observation tool)*. This variable is only defined for the teacher(s) whose lesson was observed, and is missing otherwise.

3. SSRP take-up index, based on the following variables:
    a. Indicator for teacher attending training. Set to 0.5 if teacher did not attend training but another teacher or headmaster provided instructions on how to conduct Utkarsh.
    b. Indicator for teacher having a teacher handbook. Set to 0.5 if teacher has to share.
    c. Indicator for students having student handbooks. Set to 0.5 if some students have handbooks.

4. Attendance during unannounced monitoring visit as measured by the enumerator (dummy variable).

For outcomes 1-3, control-group schools always receive a score of zero.

Note that variable e for outcome 2 is measured on the endline (rather than the midline) teacher survey, as no midline surveys were done during consolidation camp.

Note that variables h and i for outcome 2 are actually measured on the classroom observation at the midline visit, and not on a teacher survey.

All teacher-level regressions will also control for a dummy for being female, age and age squared, years of experience, years of experience squared, and a vector of dummy variables for main subject taught. For teachers without baseline data, we will use information from the midline and/or endline surveys to fill in missing values of gender, age, and years of experience.

4.2.1.3 Headmaster outcomes:

1. A dummy for whether the headmaster was present at the school during the unannounced monitoring visit.

All headmaster-level regressions will also control for a dummy for being female, age and age squared, years of experience and years of experience squared. For headmasters without baseline data, we will use information from the midline and/or endline surveys to fill in missing values of gender, age, and years of experience. We also include two school-level controls: a dummy for the school being multi-section and total enrollment.

4.2.2 Secondary Outcomes

We also analyze a set of secondary outcomes. These analyses are of interest primarily for the purpose of understanding mechanisms, or else are speculative in nature. We exclude these outcomes from our multiple testing adjustments and consider any findings on them to be exploratory rather than confirmatory.

Unless otherwise specified, all control variables for these analyses parallel those for the equivalent variable in Section 4.2.1, e.g. test score analyses control for baseline values of all three test scores, etc.

4.2.2.1 Student outcomes:

1. Standardized test score on the English, Odia, Science, and Math exams follow-up by the type of question:
    a. Foundational skills (defined as 5^{th} grade or lower competencies)
    b. PISA questions (only for English and Math)
2. Treatment effect heterogeneity for the four test scores by baseline test scores.
    a. Specifically, we will divide the baseline scores into terciles for each subject, and interact indicators for each tercile with the two treatment indicators. For science,

since there is no baseline score, we will instead use the average of the standardized baseline scores in the other three subjects.

3. Treatment effect heterogeneity for the four test scores by student gender (a dummy for the student being female)
4. Treatment effect heterogeneity for the four test scores by a dummy for the student belonging to either one of a scheduled caste or tribe (SC/ST) or an other backward caste (OBC).
5. Quantile treatment effects of each program variant on each of the four endline student exam scores. We will do this for all quintiles from the 5th to the 95th percentile.

4.2.2.3 Classroom outcomes

Our classroom outcomes are all measured at the monitoring visits:

1. Dummy for a teacher present in the classroom
2. Dummy for teaching and learning materials being visible and/or being used
3. Dummy for at least one student having an opportunity to express their own idea
4. An index based on the 2 classroom observation questions from the monitoring visit that noted whether the teacher asked a question to the class, and whether they asked a question to an individual student.
5. An index based on the 3 classroom observation questions from the monitoring visit that noted whether the teacher answered a student's question in a way that is supporting, answered in a way that is disrespectful, or ignored a question.
6. An index based on observations from monitoring visit that recorded whether the teacher seemed familiar with the content, encouraging to students, and responding satisfactorily to student questions (all measured on a 1-3 scale).
7. An index based on the student behavior from monitoring visit that recorded how many students arrived late, actively engaged in the class, not paying attention, talking to one-another about educational and non-education activities, are doing learning-related writing, and left classroom early (all measured on a 1-5 scale).
8. An index for quality of classroom teaching practice based on classroom observations, including the following measures:
    a. Dummy for teaching with student interaction (will assign 1 if the teacher is ever observed asking or responding to questions, inviting students to write on the board, or providing feedback to students and 0 otherwise)
    b. Classroom order (scale of 1-3)
    c. Teacher's position in the classroom (will assign 1 for moving around the classroom or sitting with students and 0 for sitting at own desk or at front of the room)
    d. Dummy for whether students are observed writing something

We will estimate 2 through 8 both conditional and unconditional on a teacher being present.

All control variables will be as of the baseline survey. The additional control variables for the classroom-level regressions are baseline classroom size, and controls for school level facilities (dummy variables for having a girls' toilet, a boys' toilet, being connected to the electric grid, having water for students to drink during the day, and participating in the Utthan program.

All classrooms will be included in our analysis even if the classroom teacher is not the same from the baseline or if someone other than the classroom teacher is teaching the class.

4.2.2.4 Teacher outcomes:

The outcome variables for teachers are:

1. Self-reported teacher effectiveness at endline
2. Self-reported number of days in past 6 weeks that the teacher taught an SSRP lesson at the monitoring visit
3. Whether they are currently teaching the planned SSRP lesson according to the timetable or their Flexible plan during the monitoring visit
4. Work stress/burnout index based on 5 questions that asked whether teachers report feeling mentally exhausted, fatigued, that they are positively influencing other people, energetic, and satisfied, as measured at endline.
5. Whether the teacher would want to become a teacher if they could start a new career
6. Teacher autonomy index based on the 6 questions that asked about having control over selecting learning activities, the standard of behavior, course timetable, selecting evaluation and assessment activities, selecting teaching materials, and about feeling pressure to complete curriculum, as measured at endline.
7. Stress as measured on the *Depression Anxiety Stress Scale (DASS)* module during the endline survey.
8. Anxiety as measured on the *Depression Anxiety Stress Scale (DASS)* module during the endline survey.
9. SSRP Lesson Plan Adherence index:
   a. Fraction of past 6 days that teacher followed planned Utkarsh lesson/worksheet.
   b. Indicator for the teacher conducting the correct Utkarsh lessons for this week. This is always missing for Flexible Utkarsh schools during the SLP phase.
   c. Filled out Flexible SSRP Teaching plan. This is always missing for Standard Utkarsh schools.
   d. Indicator for doing any Utkarsh worksheet (from the student handbook) on most recent teaching day.
10. SSRP Curriculum Adherence index:
    a. Share of material taught during the Foundation Camp that matches with the Utkarsh curriculum
    b. Share of material taught during the Supported Learning Phase that matches with the Utkarsh curriculum
    These are imputed to zero If FC or SSRP was not implemented when they were supposed to, and missing if FC or SLP is supposed to implemented in the future

4.2.2.5 Headmaster and School Level outcomes:

1. Headmaster's self-reported attendance as measured at the monitoring visits.
2. Headmasters' self-assessment of their effectiveness measured at the monitoring visit. We will assign a value of 5 if the headmaster considers their performance to be *much more* effective than others in similar schools, 4 if they consider their performance to be *more*

effective than others in similar schools, 3 if they assess their performance to be *just as* effective than others in similar schools, 2 if they judge their performance to be *less* effective than others in similar schools, and 1 if they consider their performance to be *much less* effective than others in similar schools and divide by 5.

3. School-level teacher attendance as measured at the monitoring visits

We will calculate school-level attendance using information from the Monitoring Visit arrival survey. We will take the total number of teachers who were supposed to be in school on the day of the visit as the denominator and the numerator will be the total number of teachers present in school during the unannounced spot check.

4.2.3 Other exploratory analyses

This document does not contain a comprehensive description of all the analyses we will conduct. We plan to do additional exploratory analyses of our data in order to understand the mechanisms behind our results.

4.3 Multiple testing adjustments

We will conduct multiple testing adjustments only for the primary outcomes specified in Section 4.2.1, since the secondary outcomes are either speculative in nature or shed light on mechanisms rather than core effects of the program. Moreover, our set of primary outcomes includes a range of independent hypothesis tests that are not simply testing the same outcome repeatedly. We therefore will only conduct multiple testing adjustments within the following families of outcomes:

Individual outcomes (no multiple testing adjustment)

The following analyses are testing distinct hypotheses and so we do not adjust their p-values for multiple testing:

- Student attendance (student outcome 5)
- Teacher attendance (teacher outcome 4)
- Headmaster attendance (headmaster outcome 1)

Student outcomes

- The four exam scores (outcomes 1-4)

Teacher outcomes

- Outcomes 1-3, which are about adherence to and implementation of the program.

To do the multiple testing adjustments, we will construct adjusted q-values using the Haushofer and Shapiro (2016) implementation of the Anderson (2008) Family-Wise Error Rate (FWER) adjustment.

Note that even for a single outcome, there are three comparisons (C vs. T1, C vs. T2, and T1 vs. T2.) So for example the exam score adjustments will be done on a set of 12 p-values.

4.4 Attrition

We define attrition as a student not having any test scores at endline. We will test for differences by study arm in the attrition rate of students who were tested in the baseline. We will also look for differences in patterns of attrition by interacting a dummy for being present at endline with a set of baseline covariates; we will use the same set of student variables from the re-randomization process described in Section 3.3 for this analysis.

We will construct Lee bounds for our test score outcomes to bound the effect of differential attrition.

We will also address differential attrition by exploiting the randomized additional mop-up visit, using the approach of DiNardo, McCrary, and Sanbonmatsu (2006) to bound the potential bias due to systematic attrition across study arms.

4.5 Covariate adjustment

To improve precision, we will estimate a variant of Equation 1, which will additionally control for covariates selected using post-double-selection LASSO method over the universe of variables in the baseline data. We will impose that our main set of controls always be selected by this method. In addition to feeding the post-double-selection LASSO algorithm linear terms in the baseline variables, we will also give it all first-order interactions between variables, and, for non-binary variables, quadratic, cubic, quartic, and logged terms. To avoid dropping zeroes we will use the inverse hyperbolic sine transform instead of the log function.

If the specification with covariate adjustment is considerably more precise (i.e. if it reduces the standard errors on any of the primary outcomes by more than 20%), we will also include these adjusted results in our main tables alongside our primary specification that does not use this additional covariate adjustment.

4.6 Cost-Effectiveness

In addition to testing the above hypotheses, we will conduct a cost-effectiveness analysis of the two treatments for increasing student test scores and attendance.

5. Data Analysis Completed Thus Far

As part of the ongoing evaluation, the PIs have access to the midline data already, and have examined some of the outcome variables from that wave of data. In particular, in order to develop a plan to minimize attrition we examined differences in student attendance by treatment group during the Foundation Camp monitoring visits (i.e. for the first 100 schools). None of the

PIs have examined any other outcomes at the time that this report was submitted, and we did not have access to any endline data when writing it.