UTKARSH PROGRAM EVALUATION

Final Project Report
August 31, 2020

Utkarsh Program Evaluation

Final Project Report
August 31, 2020

**Sabrin Beg, Arnesh Chowdhury, Soumya Das,
Anne Fitzpatrick, Jason Kerwin, Adrienne Lucas,**

Khandker Wahedur Rahman[1]

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Although India has made tremendous progress in expanding access to primary and secondary school, a large number of children who are enrolled in Class 9 lack basic foundational skills. Previous work using a randomized controlled trial in Uttar Pradesh, India, using specialized training periods, or "camps," demonstrated substantial increases in student learning for primary school students (Banerjee et al., 2017). These results suggest that remedial sessions may be a promising approach to improving overall educational attainment and transforming student outcomes for primary students. However, whether the same approach works at the secondary school level is an open question. And even if the approach is successful at the secondary level, it is unclear how to scale remedial sessions effectively: should teachers have more flexibility and autonomy in content delivery, or should lessons and content be more standardized?

The Utkarsh program consists of a series of classroom learning sessions, using existing teachers, during the school day to help prepare students in class 9 to pass their 10th board exams. 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.

This program was developed by Transform Schools, People For Action (PFA) in collaboration with the Kusuma Trust UK (KTUK) and implemented in collaboration with the Odisha Department of School and Mass Education (SME).

To rigorously evaluate this program, a randomized controlled trial (RCT) was conducted in 300 schools randomly selected from two districts of Odisha. These 300 schools were equally divided into 3 groups: 100 schools received the Utkarsh program as originally designed with a scripted lesson plan and timeline ("Standard Utkarsh"); 100 schools received the Utkarsh program with increased teacher autonomy over lesson content and delivery ("Flexible Utkarsh"); and 100 schools were the Control group and received nothing beyond the standard curriculum. The Odisha government implemented the Utkarsh program in all treatment schools.

To evaluate the effect of Utkarsh, baseline exams and surveys were conducted in all the 300 schools from July to August 2019. Abdul Latif Jameel Poverty Action Lab South Asia (J-PAL SA), study team members visited 300 schools, and collected data from a total of 299 headmasters, 831 teachers, and 5745 students for this baseline. Endline exams and surveys were conducted in all schools between December 2019 and February 2020. As a part of the endline, J-PAL SA's study team members visited 300 schools and collected data from a total of 297 headmasters, 1,121 teachers and, 5,479 students. The timeline for the endline data collection proceeded as planned, and no significant delays impacted the data collection process. In addition, monitoring visits were conducted in every school to collect information on classroom activities including teacher-student interactions, the visibility and use of teaching and learning materials, and student engagement through classroom observation. During the monitoring visits, short surveys with teachers and headmasters were conducted, and student

and teacher attendance were recorded.

Our evaluation design and instruments were based upon a pilot evaluation conducted during the 2018- 2019 academic year in which student exams, surveys, and data collection procedures were developed and validated.

We find that the Utkarsh intervention significantly improved learning for Class 9 students. Students subject to the Utkarsh treatment increased their learning in Math, English, Odia, and Science by about 0.1 standard deviations with no significant difference between the Standard and the Flexible treatments. These score increases are 40 (Odia) to 200 percent (math) more than the control group learned over this period.

The program improved student learning for all subjects. In most subjects the program benefitted students throughout the learning distribution, although in English students who were initially at lower competencies at baseline improved more. We find no effect of the program on student attendance. While we find no effect on the likelihood of taking private tuitions, students in the Standard Utkarsh program reported spending about 8 percent (Rs. 27) less on private tuitions. Students in the Standard Utkarsh program also report slightly higher aspirations of completing a bachelor's degree or higher.

While the program did not change the likelihood of teachers being present in the classroom, we find that Utkarsh did change classroom operations. We find that students are more likely to independently express an idea in Utkarsh classrooms, and that teachers in the Standard Utkarsh classroom are particularly likely to engage students. The quality of teaching practice also increased in all Utkarsh classrooms.

Section 1. INTRODUCTION

1.1 Context

Utkarsh is a government-led remedial and grade-level program for class 9 students, and focuses on 4 subjects: Odia, English, Math and Science.[2] 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 about four or more grade levels behind, and builds their foundational skills. The second phase, the Supported Learning Phase (SLP), adds medium-level students 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 grade 10 and the grade 10 board exams. These sessions operate within the typical school day with existing teachers, potentially enhancing cost-effectiveness and scalability. The program brings students to grade-level at an opportune time as students start to prepare for their board exams in class 10, a necessary hurdle to graduate from lower secondary school. Enhancing students' achievement on this high-stakes, path-defining exam can increase the likelihood of transition to higher secondary school and improve later life educational and labor-market outcomes.

This report describes the randomized evaluation of the Utkarsh Program in the state of Odisha, India.

1.2 Utkarsh Program

The Utkarsh program trained existing teachers in a new pedagogical strategy that modifies the classroom content to improve learning outcomes for students behind grade level. From the outset, the program used existing government employees.

Before the start of the academic year, education specialists from People For Action (PFA) trained district-level teams of Master Trainers (MTs). MTs are subject experts and are trained and oriented on all aspects of Utkarsh. The Master Trainers then conducted a one-day training session for the State Resource Group (SRG) on the Implementation Guide (IG) and Capacity Building Manual (CBM). SRG, created by the Department of School and Mass Education (SME), reviewed and accredited the program model, approach and resources. SRG members are subject experts and trainers from government academic institutions who monitor the progress and quality of district level training and implementation at schools and share its observations with state government. Members from the SME and the MTs oriented

---

[2] This program was previously known as the Secondary School Readiness Program (SSRP).

the District Education Officers (DEOs) on district level implementation guidelines. The DEOs and MTs then provided training and support to existing class 9 teachers and headmasters from program schools during a one-week training centered around the teacher implementation guide and how to incorporate Utkarsh subject-specific handbooks into an effective teaching practice. The headmaster and teacher training occurred right before the beginning of the academic year. PFA provided technical and financial assistance for the program and collaborated with the government to maintain quality.

At the beginning of the academic year, PFA along with assistance from SME, implemented the program in all program schools. As a part of the implementation, all program schools received baseline assessments and teaching and learning materials, e.g. Utkarsh teacher and student handbooks and workbooks from PFA.

The Utkarsh teaching and learning materials, developed by PFA, consists of tailored lessons, teaching methods, and worksheets for each learning level. First, to establish existing learning levels, schools are asked to conduct and grade levelling tests for every class 9 student. The levelling exams occur over a period of two days and cover Odia (the local language), English, Science, and Mathematics. Teachers used a rubric provided by PFA to score the tests and convert the scores into effective grade levels. After the assignment, according to their grade levels, each phase of the program was administered to the appropriate students based on their learning levels.

The lowest level learners, those who tested at the 5th-grade level or below, participated in Foundation Camp (FC) near the start of the school year. In FC, students received four hours of remedial instruction every day over 18 instructional days (a minimum of 72 hours). Students who performed above the grade 5 level, participated in alternative self-learning activities such as class 9 course curriculum, grade level lessons or higher grade concepts during the remedial classes. Once students completed FC, they were combined with students assessed to be at learning levels between grades 6 and 8 for the Supported Learning Phase (SLP). During SLP, students received three hours of remedial teaching each day for 45 instructional days (a minimum of 135 hours of teaching). Students at grade level continued with self-learning activities.

The final phase of the program was Consolidation Camp (CC), which was intended to prepare all students for the transition to grade-level material and ultimately the class 10 board exam. This phase involved 18 hours of class 9 materials over six days. At the end of the CC phase, schools received endline assessments from PFA and were asked to assess all students on their learning through these assessments. The endline assessments occured over a period of two days and covered Odia (the local language), English, Science, and Mathematics.

Section 2.   INTERVENTION AND IMPLEMENTATION

2. 1   Description of the Intervention

As programs scale up, tension emerges between harmonizing content to ensure uniform implementation and allowing flexibility to adapt to school-specific needs and conditions. On the one hand, monitoring teachers and enforcing a set schedule and lesson plan can ensure teachers follow best practices and implement programs as designed. On the other hand, teachers may be best informed on how to adapt a program to optimally meet the needs of their students and the school. This study measures the impact of implementing Utkarsh in two different ways: 1) Standard Utkarsh as fully designed, and 2) Flexible Utkarsh where teachers choose which lessons during the Learning Phase (SLP) and Consolidation Camp (CC) to adopt in their classrooms.[3]

Standard Utkarsh is a fixed curriculum for the first two learning phases, with more flexibility during CC. Teachers are provided materials for the Utkarsh lessons during the teacher training sessions and implement the program's topics and lesson plans according to the pre-specified schedule. During CC, teachers are free to choose and plan their own lessons.

Flexible Utkarsh differed from the Standard version by offering more autonomy and partial flexibility to teachers to adapt topics and/or timelines according to the needs of their students in the second and longest phase of the program, the Supported Learning Phase (SLP). During the SLP, teachers could revise their teaching plan—choosing which Utkarsh lessons to implement, in which order, and adjust time allocation for topics covered. Teachers were instructed to deliver at least 50 percent (23 days) of the total sessions provided using topics from Utkarsh teacher handbooks, but did not have to deliver those topics on the originally-scheduled day. Teachers created their schemes using the Flexible Utkarsh Teaching Plan. The Flexible Utkarsh Teaching Plan, a format created by PFA, tracked the lessons that teachers chose to implement. Teachers had the flexibility and scope to update the plans on a weekly basis based on student learning needs. The other two phases of the program (FC and CC) were implemented as usual.

We now describe each component of the intervention.

Foundation Camp (FC)

The purpose of the FC was to develop foundational concepts and skills in Odia, English, Science and Mathematics for students assessed to be below Class 5 level in the levelling test.

---

[3] All schools were instructed to implement the Foundation Camp as designed in the Standard Utkarsh program

This phase was conducted for 18 days and targeted the core competencies of students. The program consisted of 4 hours of remedial teaching every day, with a minimum time allocation of 72 hours. Head teachers reviewed the participation of teachers and students alike and planed for the next phase of the program after reviewing the academic progress.

The training for FC began with a three-day orientation of the Master Trainers (MTs)—who are subject experts—on all aspects of the program. After that, the Master Trainers conducted training sessions with various stakeholders, such as District Education Officers and the State Resource Groups. They, in turn, trained headmasters and teachers, who conduct the remedial sessions.

Both Standard and Flexible intervention schools were instructed to complete the entire FC as designed.

Supported Learning Phase (SLP)

The purpose of the Supported Learning Phase was to develop the application of foundational concepts for students assessed to be below Class 8 level in the baseline. This phase of the program consisted of 3 hours of remedial teaching every school day for 45 days, with a minimum requirement of 135 hours of teaching.

During the training for SLP, teachers in the Flexible Utkarsh arm were assisted by the trainers in completing a Flexible Utkarsh Teaching Plan in which they outlined which topics they chose to do each week of the SLP implementation. Students received the same student handbook as in the Standard Utkarsh program.

Consolidation Camp (CC)

The purpose of CC was to prepare all students for the Class 9 annual examinations. CC was organized for a minimum of 18 hours over six days for students on the syllabus for Class 9 for Odia, English, Science, and Mathematics. After this phase, the evaluation endline assessments were administered to Class 9 students to track changes in learning levels. Towards the end of the academic year, a review of student participation, endline assessments and Class 9 exam results was conducted by the headmasters to assess student learning progress.

2. 2   Collaboration and Coordination

This study was conducted with our partners, Transform Schools, People for Action (PFA) and the Department of School & Mass Education (SME), Government of Odisha. PFA, a non-governmental organization, designed all program materials.

PFA was responsible for orienting and conducting training sessions for teacher, head

teachers, District Education Officers, State Resource Group, and Master Trainers. This was done via a cascade or train-the-trainers model: a team of PFA staff who were education specialists served as core trainers, and were responsible for training Master Trainers (MTs) who in turn trained the various stakeholders. PFA along with assistance from SME supervised all the intervention training programs and managed all the logistics associated with carrying out these tasks. People for Action was also responsible for Quality Assurance (QA) visits during all phases of the program for both Standard and Flexible Utkarsh. During the QA visits, monitoring plans, a template, and online tools were used to track and report on the quality of program implementation.

SME implemented all aspects of the program with technical assistance from PFA. To provide extensive support, SME created State Research Groups (SRGs). Members from the SRGs assisted PFA during training sessions and monitored the program progress throughout the study.

All data collection for the study was conducted by Abdul Latif Jameel Poverty Action Lab South Asia (J-PAL SA).

2. 3   Key Intervention Milestones and Timeliness

Preparatory work leading to the implementation of the Utkarsh program started during the 2018-2019 academic year. The following highlights the key milestones in the preparatory work leading to the implementation activities.

To inform the design and implementation of the broader randomized evaluation, PFA in collaboration with the SME organized an orientation meeting that focused on discussing and finalizing the pilot plan. The pilot was conducted in the Cuttuck district from November 2018 to February 2019. As part of pilot evaluation activities, J-PAL enumeration teams visited 24 schools, interviewed 24 headmaster and 72 Class 9 teachers and interviewed and tested 622 Class 9 pupils in English, Mathematics, Odia, and Science. In addition to the baseline and endline activities, J-PAL teams visited all schools once as a monitoring visit. During the monitoring visits J-PAL enumerators observed the first class period of the school and documented teaching practices, teacher feedback, and what activities the students are engaged in. Additionally, focus group discussions (FGDs) of teachers and students were also conducted in two of the program schools where Utkarsh was implemented during the previous school year.[4]

In March to April 2019, People For Action (PFA) received district level school data from Department of School and Mass Education (SME). In June 2019, J-PAL finalized the list of schools for the full study. Prior to the beginning of the academic year 2019-2020, a meeting to discuss the pilot findings occurred. This included officials from SME, PFA, and J-PAL. Feedback from the piloting informed the revisions to all training, survey instruments and

---

[4] The Utkarsh program was previously known as the Secondary School Readiness Program (SSRP).

assessments for the full study. Figure 1 displays the study timeline.
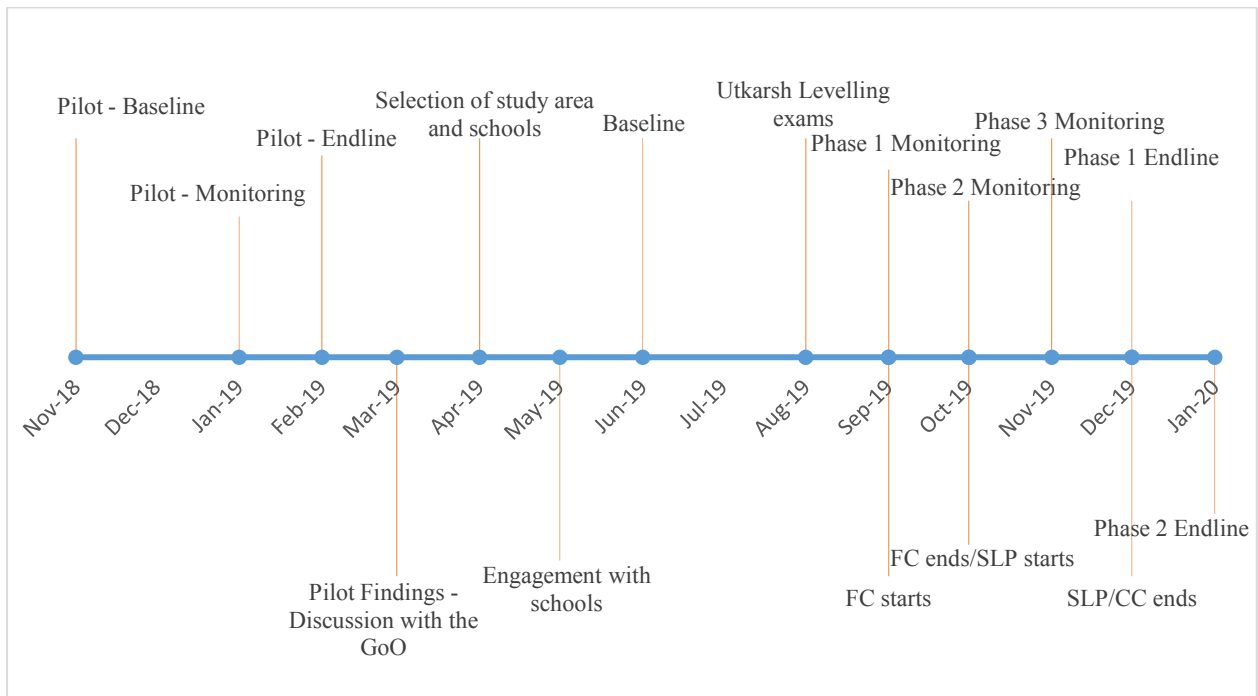
Figure 1: Study Timeline

Section 3.   EVALUATION

3. 1   Evaluation Objectives

The Utkarsh program leverages existing personnel and structures in the education sector to improve overall educational attainment. This evaluation answers the following primary research questions:

1. Does the Utkarsh program improve student learning in Math, English, Odia, and Science?

2. Which method of program implementation—Flexible or Standard Utkarsh—is the most effective at improving student learning?

3. Does the Utkarsh program improve student or teacher attendance in secondary schools, as measured by unannounced spot checks or monitoring visits?

In addition to these primary outcomes, we conducted secondary analysis to measure effects on additional outcomes and explore the mechanisms.

3. 2　Randomization Design

This randomized controlled trial operated in 300 randomly selected schools within Jajpur and Dhenkenal districts in Odisha State. The Utkarsh evaluation had three study arms: Standard Utkarsh, Flexible Utkarsh, and the control group. The 300 schools were randomly divided into three study arms. The study design is pictured in Figure 2.
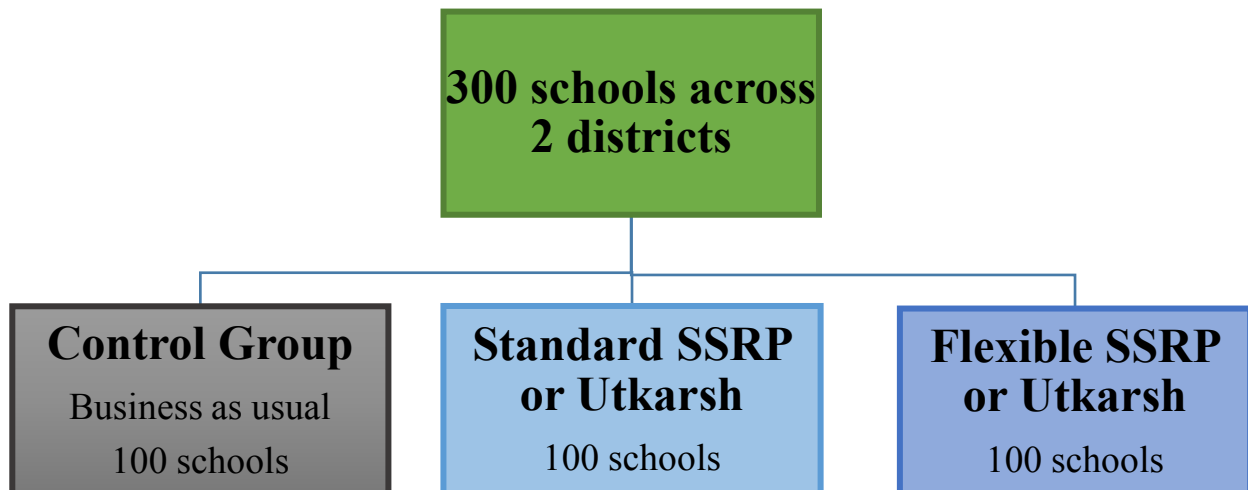
Figure 2: Randomization Design

At each school, 20 class 9 students were randomly selected to participate in the study. At each school, we interviewed 4 teachers and the headmaster or other person in-charge of the school. Due to low enrolment, absenteeism, and teacher or headmaster vacancies, the actual number of respondents per school differed slightly than what was planned.

3. 3　Evaluation Approach and Methods

3.3.1　Sample Selection

School Selection:

The school sample was constructed from the universe of all 711 government secondary schools in Jajpur and Dhenkanal. These two districts were selected after discussion with the Department of School and Mass Education (SME), Odisha.[5] In this context, it is not uncommon for headmasters and teachers to be posted in multiple schools. To reduce the likelihood of headmasters or teachers crossing treatment status, we restricted the sample to one

---

[5] Originally, Puri district was also targeted for inclusion in the Utkarsh evaluation. However, just before data collection was to begin, Cyclone Fani hit Odisha, causing extensive damage to this district. As a result of the devastation and infrastructure disruption we dropped this district from our evaluation.

(randomly selected) school per village reporting positive enrollment of class 9 students. Then, we were left with 348 schools of which we selected 300. Before a school could be included in the study, we conducted an additional screening. We excluded schools that did not use the official state language (Odia) as the instructional language, were governed by the SC-ST Development Department, did not have any enrolled class 9 students, or were schools for special needs students. Schools governed by the SC-ST Development Department have a different line of administration and do not fall within the purview of the School and Mass Education Department (SME). If a school from the list of first 300 schools did not clear the eligibility criteria, we replaced it with one of the schools from the list of the remaining 48 schools. 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, class 9 enrollment, teachers per student, Class 10 board exam passing rate in the previous year, and distance from the district headquarters.

The final sample consists of 300 schools, of which 200 schools received a version of the program (100 Standard Utkarsh and 100 Flexible Utkarsh) and the other 100 received no additional support. Due to the differing number of schools in each district, the sample was allocated proportional to each district's size.

Study Participants:

This study had three respondent types: headmasters, teachers, and students. Each respondent type completed 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 types of participants, the expected and actual sample sizes at baseline and endline, details of how they were 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 class 9 and present the day of the baseline survey visit. The universe of students was obtained from the school's student enrolment register. We first randomly selected a section (if a multi-section school) and from that section randomly selected 20 students to participate in the study. Sections and students were randomly sampled using a SurveyCTO randomization program. If a section or school had fewer than 20 students, all students were included. Our baseline sample consisted of 5,745 students.[6] We administered a short one-

---

[6] The primary reason for interviewing and assessing fewer students than 20 per school is due to low enrolment and student absenteeism resulting in fewer than 20 available students on the day of our visit. A small number of students were also excluded due to blindness, deafness, or an insufficient understanding of Odia (required in order to administer informed assent). Note that some students left midway through the school day during both the baseline and endline enumeration and so there are unequal numbers of students across the surveys and the subject exams.

on-one demographic survey, and a group assessment of their Odia, Math, and English skills.[7] During the baseline, in total 5,736 students completed the Odia exam; 5,720 students completed the Math exam; 5,721 students completed the English exam.

At the midline/monitoring visits, we collected data whether the study students were present or absent.

For the endline, we attempted to re-interview and test all sampled students from baseline, and conducted additional mop-up visits to limit attrition. Students were administered a short student survey, and then administered 4 group exams: Odia, English, Mathematics, and Science. In the endline, 5,479 students completed the Odia exam; 5,474 students completed the Math exam; 5,464 students completed the English exam; 5,466 students completed the Science exams. The sample sizes vary slightly by subject due to students leaving during the school day. For the endline analysis we used data from 5,482 students.[8]

2. **Headmasters:** In-person surveys were conducted with the headmaster or person 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 (as one Headmaster refused in the control arm). At the endline we collected data from 297 headmasters, with 98 in the control arm, 99 in the Flexible Utkarsh arm, and 100 in the Standard Utkarsh arm. For the endline survey, three Headmasters refused to participate in the survey.

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 class 9 teachers who taught Odia, English, Math, or Science in the 300 study schools were targeted for study inclusion. Our baseline sample consists of 831 teachers, or an average of 2.7 teachers per school (69% of our target). Reasons for not interviewing 4 teachers per school include absenteeism and teacher vacancies. In addition, in schools in which the headmaster was also a teacher that individual was interviewed during the headmaster survey. Below are additional details of our teacher selection procedures at baseline:

   a. For single section schools (i.e., all class 9 students were taught as one group): We aimed to interview all four subject teachers (Math, English, Odia, and Science). We selected the senior-most or the most-experienced teacher in that school for the subjects that had more than one teacher. Unavailable teachers were replaced by first selecting a teacher that matched the subject of the missing class 9 teacher. If there

---

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

[8] Not all students were reached on the day of the school visit. We attempted to reach students not present during the endline visit through mop-up visits. In total, data was collected from 983 students during the mop-up visits.

were no teachers for the same subject, then we selected other subject teachers in the following order: Math teacher, English teacher, Odia teacher, and, Science teacher. We excluded guest teachers and others who did not teach class 9 students full-time as part of our selection criteria during baseline.

b. For schools that had multiple sections: We interviewed all four subject teachers (Math, English, Odia, and Science) from the randomly sampled section. We selected the senior-most or the most experienced teacher (from the sampled section) in that school for the subjects that had more than one teacher. If teachers in the sampled section for a subject were missing or absent, we replaced them with a teacher who teaches the same subject from another section. If there were no teachers for the same subject from other sections, we selected other subject teachers in the following order: Math teacher, English teacher, Odia teacher, and, Science teacher from first the sampled section[9]; then, another section/s. We excluded guest teachers and other teachers who do not teach class 9 students full time as part of our selection criteria during baseline.

In this setting, creating a consistent and comparable sample of teachers across the study waves is challenging due to teacher vacancies and absenteeism and guest teachers (who may not work at the school each day of the week). The teachers surveyed at midline and endline were therefore not always the same as those surveyed at baseline because of turnover, because some teachers were missed at baseline, and because we allowed teachers who taught in the school part-time to be included in our study if there were an insufficient number of teacher surveys in the school. To reduce respondent burden, we also created a modified teacher/headmaster tool for schools in which a selected teacher also performed the duties of the headmaster. Finally, for the midline and endline, we deemed it of higher priority to interview the teacher implementing the Utkarsh lesson(s), who may or may not have been one of the teachers who were interviewed at baseline. Below we describe our teacher selection procedures for the endline.

For the endline, four class 9 teachers at each school were targeted for inclusion in the endline data collection. In total, we conducted surveys with 1,121 teachers, an average of 3.7 teachers per school. Reasons for being unable to interview 4 teachers per school include absenteeism, understaffed schools, schools in which the headmaster is also a teacher (and therefore that individual was interviewed during the headmaster survey). For the endline, we surveyed more teachers than we did for baseline, because teacher(s) who did not clear our endline teacher selection criteria but had participated in baseline and monitoring visits were surveyed during mop-up visits to fulfil the target of 4 teachers per school.

a. For treatment schools (i.e., those schools assigned to Standard as well as Flexible Utkarsh): Priority was given to teachers teaching the Utkarsh subjects—Math, English, Odia, and Science. One Utkarsh teacher would be

---

[9] Some schools had no English teacher but multiple Math teachers, for example.

selected for each of the 4 assigned subjects, i.e., Math, English, Odia and Science. If there were multiple Utkarsh teachers for a given subject, we interviewed the teacher who mostly taught the students in the sampled section of class 9. If the designated Utkarsh teachers were not available for surveys from any of the four subjects, we interviewed a different Utkarsh teacher. If Utkarsh teachers for a subject were missing or absent, we replaced the teacher with another Utkarsh teacher in the order of Math, English, Odia and lastly Science with the following criteria: (i) the sampled section, (ii) any other class 9 section (iii) a teacher from a previous survey round (first monitoring, and if unavailable from monitoring, then baseline). In instances in which the headmaster or person in-charge was also a teacher for one of our four target subjects, we had that respondent complete an abridged version of both surveys. We replaced missing teachers with any non-Utkarsh teacher, either from the same or from a different subject, in case there were still fewer than four teacher surveys.

b. For control schools: Priority was to interview the Class 9 teachers for Math, English, Odia, and Science from the stream sampled at baseline. If any teacher was not available during the day of data collection, we replaced the teacher in the following order: (i) another teacher for the same subject and the same stream, (ii) another teacher for the same subject from another stream (iii) another Math, English, Odia, or Science teacher from the same or different stream; (iv) a teacher from baseline. In the case of overlapping criteria (if two teachers could be selected due to identical fulfilment of criteria), we selected the most-senior teachers for the survey and proceeded in the order of seniority. We replaced missing teachers with any other subject teacher if there were less than four teacher surveys. In instances in which the headmaster or person in-charge was also a teacher for one of our four target subjects, we had that respondent complete an abridged version of both surveys.

3.3.3. Research Instruments

Our baseline and endline data collection involved surveys of three respondent types: (1) headmasters, (2) teachers, and (3) students. We used 9 survey instruments during the endline surveys, of which 6 instruments were similar to the instruments from the baseline. The 6 main tools included the Arrival Survey, Headmaster Survey, Teacher Survey, Student Survey, Student Assessment Tests, and the Roll Call instrument. The questionnaire format followed that of the baseline. In addition to the main survey instruments used from the baseline, we used 3 additional instruments for the endline data collection, namely Student Board Registration, Enumerator Survey, and a Teacher Competency Instrument.

For the monitoring visits, we used surveys of two respondent types: (1) headmaster and (2) teachers. In addition to these, we used three different surveys – Student Roll Call, Arrival, and Classroom Observation.

The Principal Investigators designed all the tools. The bulk of survey questions were drafted by the PIs specifically for this project. Some survey questions were adapted from other large-scale education evaluations, such as Beg et al. (2020) in Ghana and Teaching at the Right Level studies in three Indian states (Banerjee et al. 2017). The principal autonomy module was based upon the US Department of Education's National Teacher and Principal Survey and the teacher autonomy module was based upon the Teacher Autonomy Survey, modified to reflect the Indian context. Teacher stress questions were adapted from a subset of the Maslach Burnout Inventory. Student engagement questions are adapted from the Programme for International Student Assessment (PISA). Questions were also sourced questions from the J-PAL survey question bank and other standardized exams for secondary students. The draft surveys went through a series of iterations before they were finalized. This involved pilot-testing the surveys and pre-testing the surveys to ensure questions were clearly worded, with a clear translation, and the survey could be completed in a reasonable amount of time. These activities generated iterative feedback for the PIs, who then revised the tools at various steps before finalizing them. The teacher competency instrument was adapted from the Service Delivery Indicators (SDI) initiative of the World Bank, which instructed teachers to grade a hypothetical pupil's responses on a numeracy and literacy test.

All surveys were administered in Odia. After finalizing the instruments, the surveys were programmed in SurveyCTO. All programmed surveys were bench-tested to ensure that all consistency checks and skip patterns worked as expected.

The following presents the contents of each of the survey instruments:

a. *Arrival Survey:* The Arrival Survey was administered to any staff who had the authority to grant permission to enter the school and conduct the baseline activities (typically the headmaster, the headmaster in-charge, or another informed deputy). It also contained a teacher roster and a brief set of questions on which teachers were teaching which subjects.

b. *Headmaster Survey:* The Headmaster Survey was administered to the headmaster or the person in-charge in the sampled schools through an in-person interview. The Headmaster Survey sought information on the demographic characteristics, previous education/experience, daily activities, class 9 teacher details, general school, and student practices, including enrolment, perceptions about teaching and learning in schools, work-related stress, and autonomy. The Headmaster Survey used for monitoring visits included demographic characteristics, previous education/experience, school management practices, and perception and implementation of intervention program. In a large number of schools the headmaster position is vacant, resulting in teachers performing the duties of headmaster. Therefore, we also developed a survey which combined key questions from the teacher and headmaster instrument to reduce respondent burden from sitting through two interviews.

c. *Teacher Survey:* The Teacher Survey was administered to selected Class 9 teachers through an in-person interview. The Teacher Survey included demographic

characteristics, previous education/experience, activities conducted by them during the school day and outside, perceptions about teaching and learning in schools, work-related stress, burnout, and teacher autonomy. The Teacher Survey used for monitoring visits included demographic characteristics, previous education/experience, teaching practices, and perception and implementation of intervention program.

d. *Student Survey:* The Student Survey was administered to Class 9 students who were part of the baseline survey through an in-person interview with an interviewer of the same gender. The survey captured information on the students' aspirations and students' feedback on classroom teaching.

e. *Student Assessment for Mathematics, English, Odia, and Science:* The Student Assessments were administered to the sampled class 9 students as a group test. These instruments were adapted from a combination of resources, including the Utkarsh assessment tools and handbooks, PISA questions and resources, and other standardized exams for secondary school students. For the endline, an additional assessment for Science was done, which was not conducted during the baseline.

f. *Student Board Registration:* The student board registration form was administered to the headmaster or the person-in-charge. It collected the students' class 10 Board registration unique number and date of birth of the 20 sampled students from the class 10 board registration sheets.

g. *Student Roll Call:* The student roll call was administered in the sampled class 9 classroom. It recorded whether or not the sampled students from baseline were present in the classroom on the day of the baseline, monitoring, endline, and mop-up visits. If students were in different sections (other than the sampled section from baseline), we approached all the possible teachers of these sections and subsequently administered the roll call roster.

h. *Classroom Observation:* The Classroom Observation tool was administered for one class period inside the sampled section or classroom. It recorded classroom settings, processes, teaching practices, teacher-child interaction, and student behavior through multiple snapshots in a period.

i. *Teacher Competency Test:* The teacher competency test was administered during the endline and mop-up visits to all the sampled class 9 teachers from a subset of schools. This test was administered right after the teacher survey and required the teachers to grade a hypothetical student's responses on English and Maths papers. This tool was created and validated by researchers at the World Bank.

j. *Enumerator Survey:* The enumerator survey was administered to the enumerators through an in-person interview. The survey included questions that recorded the background characteristics of the enumerators and their work experience history.

Additionally, we used three backcheck survey instruments during the baseline and endline to ensure the quality of the responses received. The backcheck surveys were administered to three respondent types: (1) headmasters, (2) teachers, and (3) students. The backcheck data was used to identify and rectify surveyor errors and was used internally for quality assurance and monitoring purposes.

The backcheck surveys were drafted using a set of questions from the original questionnaires and were matched to the responses collected earlier. The surveys were administered to 22.6% (totalling 68 schools) and 19.6 % (totalling 59 schools) of the sampled schools during baseline and endline respectively.

3.3.4. Ethical Considerations

IFMR Institutional Review Board approved the protocols and questionnaires. Individual PIs also received approval through their home institutions. We obtained individual consent, or assent for those below the age of consent, from each headmaster, teacher, and student, at the beginning of the surveys for each wave of the study, including the monitoring visits. We obtained consent to conduct surveys, observe classrooms, and collect students' class 10 Board registration information from school headmasters. All respondents were informed of the voluntary nature of participation and the confidentiality and anonymity of information. Each participating headmaster and teacher received a small gift in consideration of their time. This gift was either a pen or a diary/book. Each study student received a light snack to ensure that they did not lose focus during the long hours of assessments.

3.3.5. Field Staff Recruitment and Training

Field staff were selected for training following a competitive recruitment process that brought together a combination of J-PAL-experienced and newly hired enumerators. Different categories of field staff were selected and trained. These were enumerators, supervisors, monitors, project associates, and back-checkers. The training sessions were primarily facilitated by Project Associates, Survey Coordinator, and the Research Associate.

Different training sessions relating to Surveyor Training, Supervisor Training, and General Training were conducted for each phase. Surveyor training sessions were directed primarily towards surveyors and highlighted the roles and responsibilities of a surveyor. Supervisor training sessions were directed towards potential candidates, i.e., candidates having the caliber of a supervisor. The roles and responsibilities of a supervisor were highlighted during these sessions. General training covered other aspects such as project description, J-PAL protocols, and field protocols. The training approach was based on classroom teaching. Classroom teaching practices comprised presentations, questions and answers, group discussions, role-plays, and mock practice sessions. A one-day monitor training was held for 6 pre-selected candidates for the position of field supervisor. This was to help them have a general overview of the surveys and conduct specialized training to enhance their participation and support to the training facilitators. The specialized training included sessions on effective ways to conduct training programs, how to score/evaluate surveyors/supervisors, and instruments' details to clarify doubts before launching the instruments in the surveyor/supervisor training programs.

For Baseline:

The baseline training was held from 6th to 22nd June 2019 (i.e., 12 days of training excluding public holidays) in multiple rented venues in Bhubaneswar. 114 people were trained on the baseline surveys and assessments. Only 75 surveyors and 15 supervisors were retained for the baseline. Each Project Associate was responsible for one district i.e., nine field teams and six field teams in Jajpur and Dhenkanal district, respectively. Every monitor (a total of 6), was responsible for three survey field teams. The survey field team comprised 90 field staff members. The entire field team consisted of 15 survey field teams with each team composed of one supervisor and five surveyors. Additionally, each district had one team of two back-checkers each who were managed and monitored by the respective PAs. The staff allocation is pictured in Figure 3.
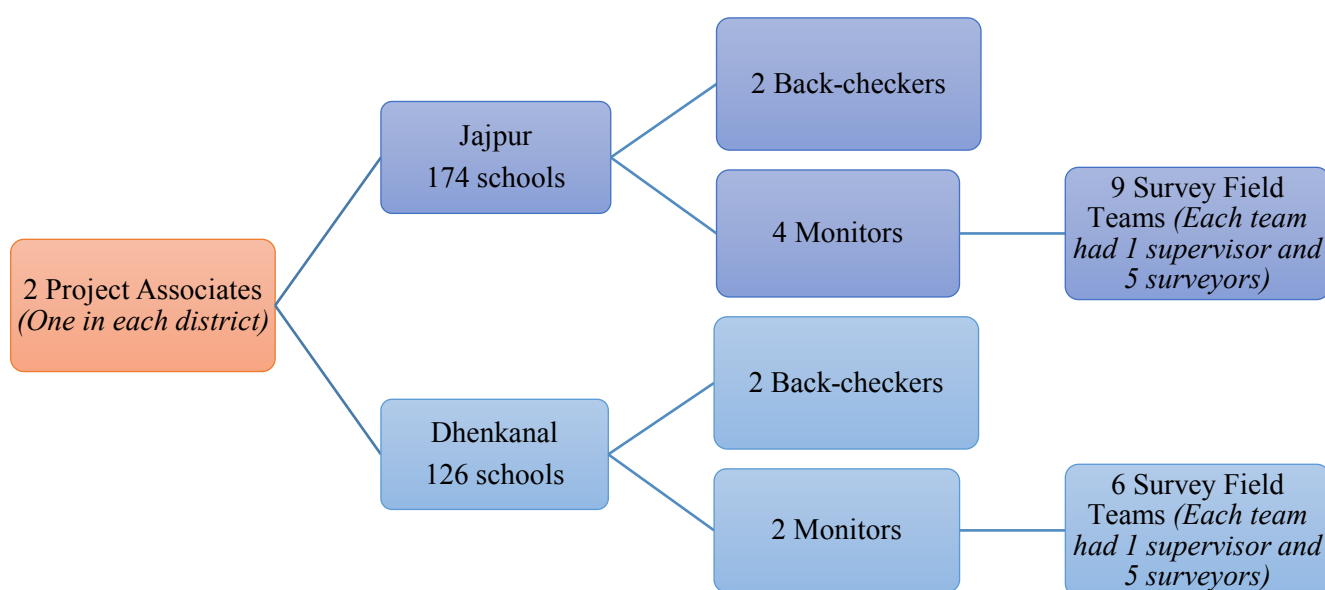
Figure 3: Staff Allocation

For Endline:

The endline was conducted in two phases to accommodate school holiday and still complete all data collection prior to the February, when schools begin preparing for end-of-year exams.

A total of 120 and 150 people were trained on the endline phase 1 and 2 surveys and assessments. Of which 77 and 125 members qualified for the fieldwork and were subsequently selected for endline phase 1 and phase 2, respectively.

The endline training for phase 1 was held from 27 November to 7 December 2019 (i.e., 10 days of training excluding public holidays). The endline training for phase 2 was held from 19 December 2019 to 4 January 2020 (i.e., 13 days of training excluding public holidays). The training sessions were held in multiple rented venues in Bhubaneswar.

A three-day monitor training was held for 7 pre-selected candidates for the position of a field supervisor. This was to help them have a general overview of the entire endline phase and enhance their participation and support to the training facilitators. Two Project Associates,

Survey Coordinator, and the Research Associate were primarily facilitating the training sessions.

Field staff structure for endline phase 1 is described below:

Both the Project Associates were allocated to the Dhenkanal district. A total of 5 monitors was responsible for at least two survey field teams. The survey field team comprised 70 field staff members. The entire field team consisted of 10 survey field teams, with each team composed of one supervisor and five surveyors. Additionally, they had two back-checkers and six mop-up surveyors who were managed and monitored by the renaming two monitors. Staff allocation for endline phase 1 is illustrated in Figure 4.
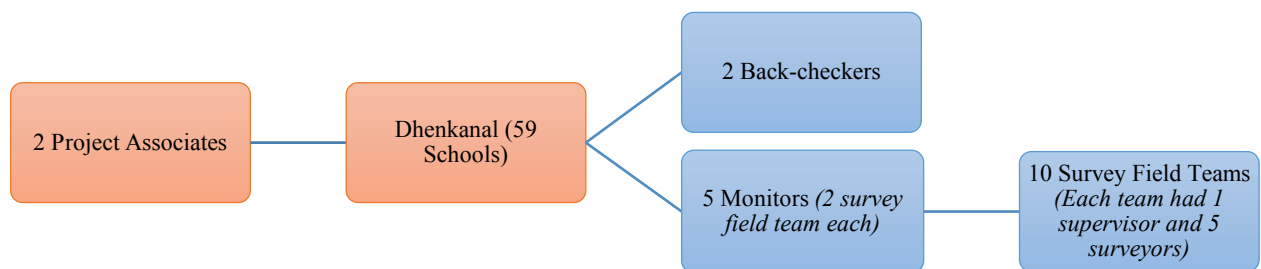
Figure 4: Staff Allocation for Endline Phase 1

Field staff structure for endline phase 2 is described below:

Each Project Associate was responsible for one district, i.e., 12 field teams and four field teams in Jajpur and Dhenkanal district, respectively. While two monitors in Dhenkanal were responsible for two survey field teams each, four monitors in Jajpur were responsible for three survey field teams each. The survey field team comprised 96 field staff members. The entire field team consisted of 16 survey field teams, with each team composed of one supervisor and five surveyors. Further, Jajpur had six mop-up teams (two surveyors each), and Dhenkanal had three mop-up teams (two surveyors each) that were managed by the PAs. Additionally, each district had two teams of two back-checkers, each who were controlled and monitored by the respective PAs. Staff allocation for endline phase 2 is illustrated in Figure 5.
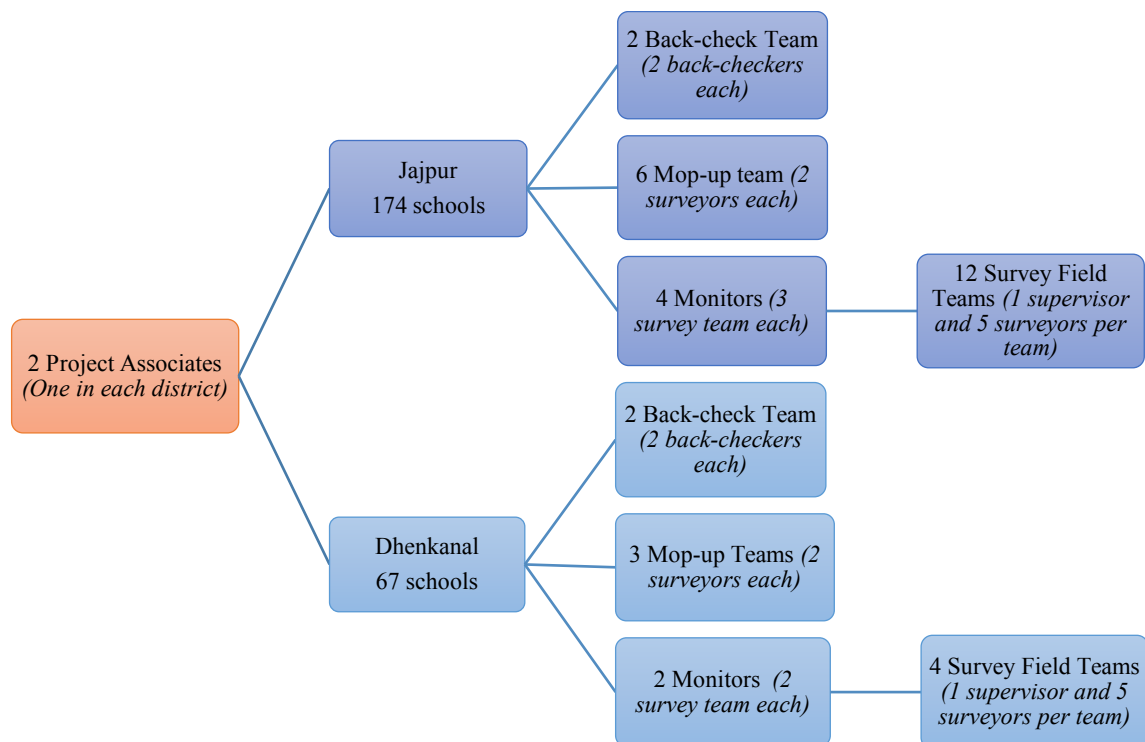
Figure 5: Staff Allocation for Endline Phase 2

3.3.6. Data Collection

We conducted 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.

Figure 1 above contains these collection activities along with other important milestones.

The endline surveys took place from December 2019 to February 2020, but school holidays (winter break) were from late December until early January. Thus, we randomly selected 60 schools to do the endline prior to the holidays, with the remaining 240 taking place afterward.[10] Prior to the holidays, we completed the endline data collection in 59 schools, and the remaining 241 schools were visited between January 2020 and February 2020.

---

[10] 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 maximize response rate, mop-up visits were conducted at schools to survey respondents (student, teacher, and headmaster) who were absent during the endline visit. In instances where the student board registration survey was incomplete during the endline visit, the board registration survey was administered to the headmaster or the headmaster in-charge during the mop-up visits. Reason for incomplete board registration survey during the endline visits were primarily because the headmaster or person in-charge did not have time to complete the survey or was not able to access digital copies of the class 10 board registration. The mop-up survey tools (student, teacher, headmaster, student board registration) were the same as the endline tools. We conducted second mop-up visits in a random subset of schools where respondents remained absent during the first mop-up visit. Seventy-six schools did not require a second visit as those schools had a 100% coverage rate by the end of the first mop-up visit.

3.3.7. Coverage

Table 1 presents the coverage rates for the instruments administered to the headmasters, teachers, and students. Student completion rates for both baseline and endline visits were less than the target rates because (a) some schools lacked 20 present Class 9 students, (b) some students did not understand Odia language, and (c) some students were special needs children. Teacher completion rates were lower than anticipated at baseline because (a) some schools lacked teachers in class 9, (b) teacher absenteeism at some schools, and (c) some schools had guest teachers as class 9 teachers. We excluded guest teachers during the teacher selection procedure in all the schools at baseline. Hence, schools having guest teachers as class 9 teachers was one of the primary contributors to low teacher survey completion rates, and thus we modified protocol for subsequent rounds.

Table 1: Coverage by Instrument

Instruments	Target	Baseline	Endline Complete	Refused	% Coverage
Headmaster Survey	300	299	297	2	99.3%
Teacher Survey	1144	831	1121	0	134.9%
Student Survey	6000	5745	5482	0	95.4%
Pupil Assessment					
English	6000	5721	5463	0	95.5%
Odia	6000	5736	5479	0	95.5%
Math	6000	5720	5474	0	95.7%
Science	6000	N/A	5466	0	N/A
Teacher Competency Test	1144	N/A	818	21	N/A

Notes: % Coverage is the fraction of baseline respondents who were found for endline. It is not applicable for instruments with no baseline. For some schools the coverage rate is more than 100% because during the endline we administered the teacher survey on teachers who were present during the day of the visit (as per the endline prioritization criteria) but then during mop-up visits we then sought to administer the survey to respondents from the baseline or monitoring visits. We did not conduct science assessment for pupils and teacher competency assessment in the baseline.

3.3.8. Data Collection Challenges

The research team encountered the following challenges.

a. Limited or no school resources to use for evaluation activities. The majority of the schools had limited extra classrooms, chairs, and tables, which hindered the setting up of a conducive environment for the interviews or pupil assessment activities.

b. Limited enrolment and pupil absenteeism in some schools. Some schools had fewer than 15 students per class, and hence the team could not meet the target of 20 students per school. In addition, data collection activities were hindered by the pupil absenteeism. High student absenteeism was observed in schools that were either preparing or organizing local festivals or sports events.

c. Accounting for schools with students who did not read, write, or speak Odia language. A few schools in Jajpur were in areas where the majority of the population speaks Urdu as their first language. We handled this issue by dropping and replacing schools where a majority of students were Urdu speakers based on an initial phone screening of schools.

Furthermore, during baseline data collection we also encountered one Urdu-language school that was not successfully screened out of our sample. We dropped this school and replaced it with a school from the backup list of schools.

d. After arriving to conduct baseline activities, we encountered a school that came under the jurisdiction of the SC-ST Development Department and consequently had to drop the school from our sample. The school was replaced with another school from the backup school list.

e. In the initial days of the endline data collection, while booking the school visit dates in the Jajpur district, we came across schools that refused to participate in our activities as they claimed to have sports tournaments on those days. Further, during the endline visits, our field team faced a lot of resistance and non-compliance from teachers and headmasters of the Jajpur district in the initial days of data collection. This resulted in creating multiple field plans, which in turn resulted in calling the headmasters and rescheduling the school visit dates frequently.

f. Teachers were reluctant to participate in the teacher competency tests. We encountered refusals and non-compliance with the teacher competency test instructions from teachers across districts. They feared they would lose their jobs if they underperformed. Our protocols were designed to assure teachers that personally identifiable data will not be disclosed to any parties. About 21 teachers refused to take the tests, and of the 802 teacher who took the tests, about 37 teachers partially

completed the tests.

g. Poor network connectivity. There was poor network connectivity due to the remote nature of the localities in which the sampled schools were located. This sometimes hindered communication among the team members and their respective monitors.

3.3.9. Baseline Balance

Despite these challenges, our team collected high quality data from all respondents. Table 2 shows that randomization successfully created three groups with balanced characteristics at baseline.

<div align="center">

Table 2: Balance Table

</div>

	Control (N=1949)	Standard Utkarsh (N=1876)	Flexible Utkarsh (N=1931)	p-value from test of equality (1)=(2)=(3)
Panel A: Student-Level Variables (N=5756)	(1)	(2)	(3)	(4)
Female Student	0.495	0.490	0.511	0.610
Age (in years)	13.163	13.148	13.152	0.875
Scheduled Caste or Scheduled Tribe (=1)	0.235	0.244	0.248	0.770
Participated in Utthan (=1)	0.878	0.892	0.879	0.395
Takes Private Tuition (=1)	0.733	0.714	0.731	0.712
	(N=276)	(N=277)	(N=278)	(1)=(2)=(3)
Panel B: Teacher-Level Variables (N=831)	(1)	(2)	(3)	(4)
Female (=1)	0.486	0.466	0.496	0.847
Age of the teacher (in years)	41.830	43.076	41.216	0.230
Have a teaching certificate (=1)	0.341	0.361	0.353	0.811
Teaching experience (years)	16.286	17.773	15.629	0.115
	(N=99)	(N=100)	(N=100)	(1)=(2)=(3)
Panel B: School-Level Variables (N=299)	(1)	(2)	(3)	(4)
Female (=1)	0.263	0.2	0.19	0.460
Designation in School: Headmaster (=1)	0.172	0.23	0.22	0.569
Sanctioned class 9 teacher posts in the school	7.545	7.83	7.38	0.310
Number of teacher posts filled	5.192	4.98	5.16	0.570
Total enrollment in class 9	72.384	62.51	72.87	0.017**

Notes: Table is showing reported characteristics of the respondents during the baseline survey. Standard errors are clustered at school. ***, **, and * indicate significance at the 1, 5, and 10 percent levels.

Panel A describes student characteristics. Approximately half of the students in the sample are girls, and the average age is approximately 13 years. Nearly one-quarter (24%) of students are scheduled caste (SC) or scheduled tribe (ST). Prior to our evaluation of Utkarsh, the Utthan program, a remedial program that focused on foundational skills for Class 8 students, was implemented in the study area. Overall, 88% of students reported participation in Utthan. Finally, about 72% of the student sample reported currently paying for private tuitions.

Panel B describes teacher characteristics. Slightly less than half of the teachers are female (48%), and the average teacher is 42 years old. Approximately one-third (35%) of teachers have a teaching certificate, and the average teacher has 16 total years of experience.

Panel C presents headmaster and school-level characteristics from the survey of headmasters (or another informed deputy if no headmaster was available). Approximately 22% of these individuals were women. In total 19% of the baseline surveys were conducted by headmasters, reflecting that many schools had vacancies for that position. Similarly, the average school has slightly more than 7 Class 9 posts, but only 5 posts were filled. Total student enrolment in Class 9 was approximately 69 students.

Across all measures, only one—total enrolment—is statistically different across the three arms. Class 9 enrolment in Standard Utkarsh schools is smaller than in the other two arms. As we are testing 13 measures, some degree of imbalance is expected and is not expected to pose a challenge to our evaluation.

Section 4.   EMPIRICAL STRATEGY

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

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

where Y_{isj} is the outcome of interest measured at endline for respondent i in school s in stratification cell j. *StandardUtkarsh* and *FlexibleUtkarsh* 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).[11] 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 are clustered at the school level.

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

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

---

[11] In all our specifications, if a control variable is missing, 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.

Tests of H01 and H02 correspond to whether β_1 and β_2 are statistically different from 0. Tests of H03 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.

Section 5. EVALUATION FINDINGS

5.1 Student Test Scores, Attrition, and Attendance

In Table 3 we present the results of the estimate of the effect of Utkarsh on student test scores.

Table 3: Test Scores

	(1)	(2)	(3)	(4)	(5)
	Overall	English	Math	Odia	Science
Standard Utkarsh	0.107***	0.118***	0.119***	0.089***	0.104***
	(0.015)	(0.020)	(0.020)	(0.017)	(0.031)
Relative to control group learning	0.570	0.557	1.887	0.424	0.556
Years of schooling (Evans and Yuan)	0.503	0.555	0.559	0.418	0.489
Flexible Utkarsh	0.110***	0.116***	0.123***	0.086***	0.143***
	(0.015)	(0.018)	(0.021)	(0.017)	(0.032)
Relative to control group learning	0.587	0.547	1.943	0.413	0.765
Years of schooling (Evans and Yuan)	0.517	0.545	0.578	0.404	0.672
Observations	5,448	5,448	5,448	5,448	5,448
Control-group change (baseline to endline)	0.187	0.212	0.063	0.209	see notes
Flexible Utkarsh=Standard Utkarsh (p value)	0.805	0.212	0.851	0.885	0.210

Notes: All regressions include strata, week, and day-of-week fixed effects; standardized IRT scores from baseline English, Math, and Odia tests, a dummy for being female, age of the pupil, and indicator variables for endline interview phase. Heteroskedasticity-robust standard errors, clustered by school, in parentheses. Columns 1-4: Relative to control group learning measures are the treatment effects divided by the average learning gain from the control group. Column 5: Because we have no baseline scores for the science, we divide by the average control group learning from column 1 (0.187). *** p<0.01 ** p<0.05 * p<0.1

Both models of the Utkarsh program increased students' test scores by about 0.11 standard deviations (SD) overall (column 1). When disaggregated by subject, scores increased about 0.12 SD for both arms for both Math and English (columns 2 and 3) and about 0.09 SD in Odia (column 4). For the Science subject score, Standard Utkarsh increased scores by 0.10 SD and the Flexible version increased scores by 0.14 SD (column 5).

To help in understanding these magnitudes, we calculated the treatment group learning relative to the control group change in learning over this same period. The changes in the control group test scores appear near the bottom of the table. The increases in the overall and English scores due to Utkarsh are about 0.5 to 0.6 beyond what the control group learned

(columns 1 and 2). Therefore, Utkarsh students learn about 50 to 60 percent more during the same period as the control group. For math, the Utkarsh students learned almost 2 times more than the control group (column 3). For Odia, the increase in test scores from Utraksh is about 40 percent more than the control group learned. We change the method slightly for the science test as students did not take a Science test at baseline. For the science test, we scale the test score using the overall score (column 1) scaling factor, finding increases of 60 (standard) to 80 (flexible) percent beyond the control group learning in Science due to Utkarsh.

As an alternative way of understanding the learning gains caused by the program, we also present the results scaled by a factor recommended by Evans and Yuan (2019). They show that each standard deviation increase in learning is associated with between 4.7 and 6.8 years of school. This conversion is based on an analysis of the literacy scores of adults from a set of low- and middle-income countries who have varying levels of formal schooling. We use their lower-bound number (4.7 years per standard deviation) and multiply it by the estimated treatment effect in terms of standard deviations. The results are mostly very similar to the figures that express the treatment effects relative to control-group learning. The one exception is for math, because our study's control group learned very little in that subject; the Evans and Yuan conversion does not differentiate by subject and so it does not take that into account.

To visualize the size of the treatment effects, we present the learning gains for each subject in Figure 7. Panel A shows how much students learned since baseline in English, Math, and Odia for each study arm. The gains are measured in standard deviations of the baseline distribution. We can see that the control group improved their scores in English and Odia by about 0.2 SDs from baseline to endline, but the treatment arms improved their scores by over 0.3 SDs in English and by nearly 0.3 SDs in Odia, a statistically-significant gap. In Math, the control group learned less, improving their scores by just over 0.05SDs. The treatment caused students to improve faster: students in both treatment arms improved their test scores by about .175 SDs, and these treatment effects are statistically significant.

We cannot show learning gains for Science because there was no baseline Science test. Instead, Panel B of Figure 7 presents the average Science test scores at endline for all three study arms. The control-group average score is standardized to zero, and the other averages are expressed in terms of standard deviations of the control-group endline distribution. These treatment effects for this subject are also statistically significant. The Standard Utkarsh program increased learning in Science by about 0.075 SDs, and the Flexible Utkarsh program by about 0.16 SDs. While the latter effect is larger, the difference between the two is not statistically-significant and thus could have occurred by chance even if the programs are equally effective.

Figure 7: Treatment Effects by Subject
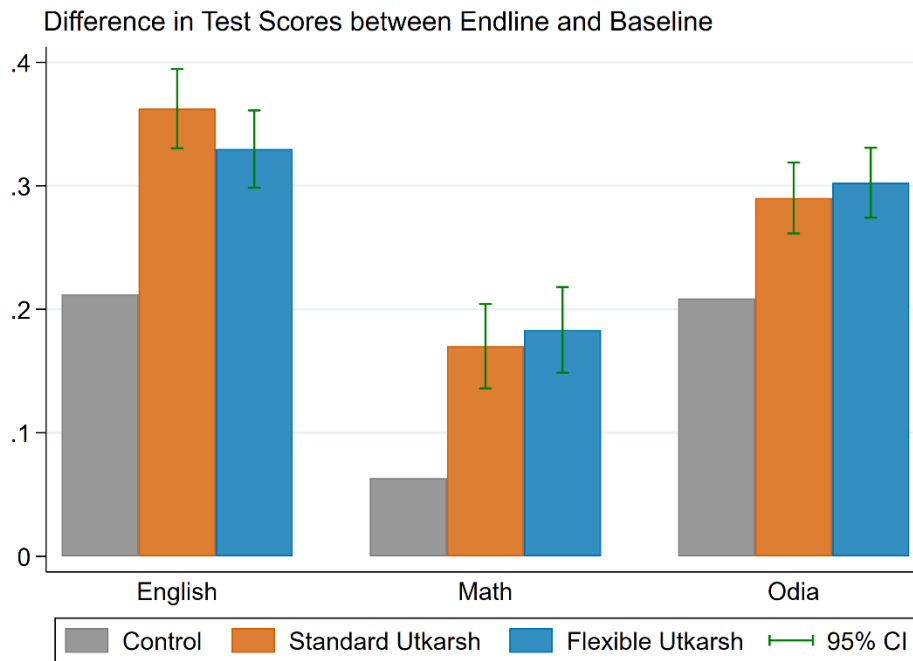
Panel A: Learning Gains (English, Math, and Odia)
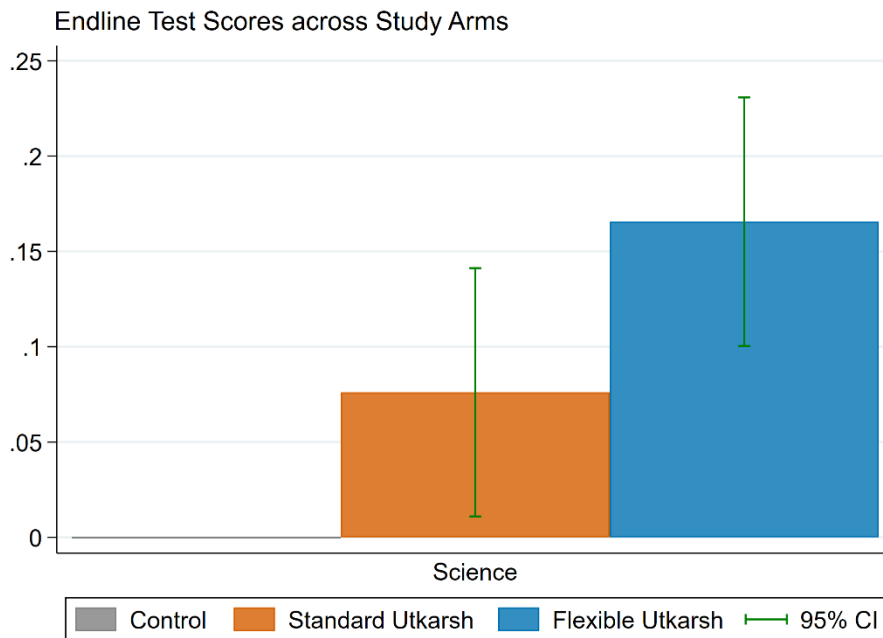
Panel B: Endline Test Scores (Science)

Notes: We cannot show the change from baseline for science as there was no baseline science test. Instead, Panel B presents the endline test scores for science. The control group test score is standardized to zero, so no bar appears for the science scores in Panel B.

We further tested for whether this effect is larger for students initially at different learning levels by testing for heterogeneity by baseline test score. We divided students into terciles (i.e., low/medium/high learners) based on their baseline test score. For simplicity and given the similarity of results between the Standard and Flexible Utkarsh above, we combined the two treatments into a single Utkarsh indicator. The results of the tests for heterogeneity appear in Table 4.

Table 4: Heterogeneous Effects on Test Scores by Baseline Test Score

	(1)	(2)	(3)	(4)	(5)
	Overall	English	Math	Odia	Science
Utkarsh*BL Score Low	0.126***	0.145***	0.131***	0.123***	0.075
	(0.024)	(0.030)	(0.033)	(0.027)	(0.046)
Utkarsh*BL Score Medium	0.102***	0.125***	0.124***	0.069***	0.108***
	(0.017)	(0.026)	(0.024)	(0.020)	(0.039)
Utkarsh*BL Score High	0.095***	0.075***	0.104***	0.071***	0.184***
	(0.015)	(0.019)	(0.022)	(0.022)	(0.038)
Observations	5,443	5,443	5,443	5,443	5,443
Control-group change (baseline to endline)	0.187	0.212	0.063	0.209	see notes
Utkarsh*BL Score Low = Utkarsh*BL Score Medium = Utkarsh*BL Score High (p value)	0.478	0.087	0.700	0.166	0.137

Notes: All regressions include strata, week, and day-of-week fixed effects; standardized IRT scores from baseline English, Math, and Odia tests, indicators for overall BL score (low/medium/high), a dummy for being female, age of the pupil, and indicator variables for endline interview phase. Heteroskedasticity-robust standard errors, clustered by school, in parentheses. Column 5: Students did not take a baseline science test. *** $p<0.01$ ** $p<0.05$ * $p<0.1$

The Utkarsh program raised student test scores across the entire distribution of initial student performance. Test scores increased for students of all terciles of overall performance across the overall, English, Math, and Odia tests. However, for Science, students in the lowest tercile did not have statistically-significant score improvements. Even though the point estimates across the different terciles are different for the Overall, Math, Odia, and Science scores, we fail to reject that they are statistically the same. For English, we reject that the scores are equivalent (p-value less than 0.10) with the largest effect sizes happening for the lower two terciles. The treatment effects on overall test scores by tercile are presented in Figure 6.

Change in Endline Test Score of Utkarsh Participants
(Standard Deviation)
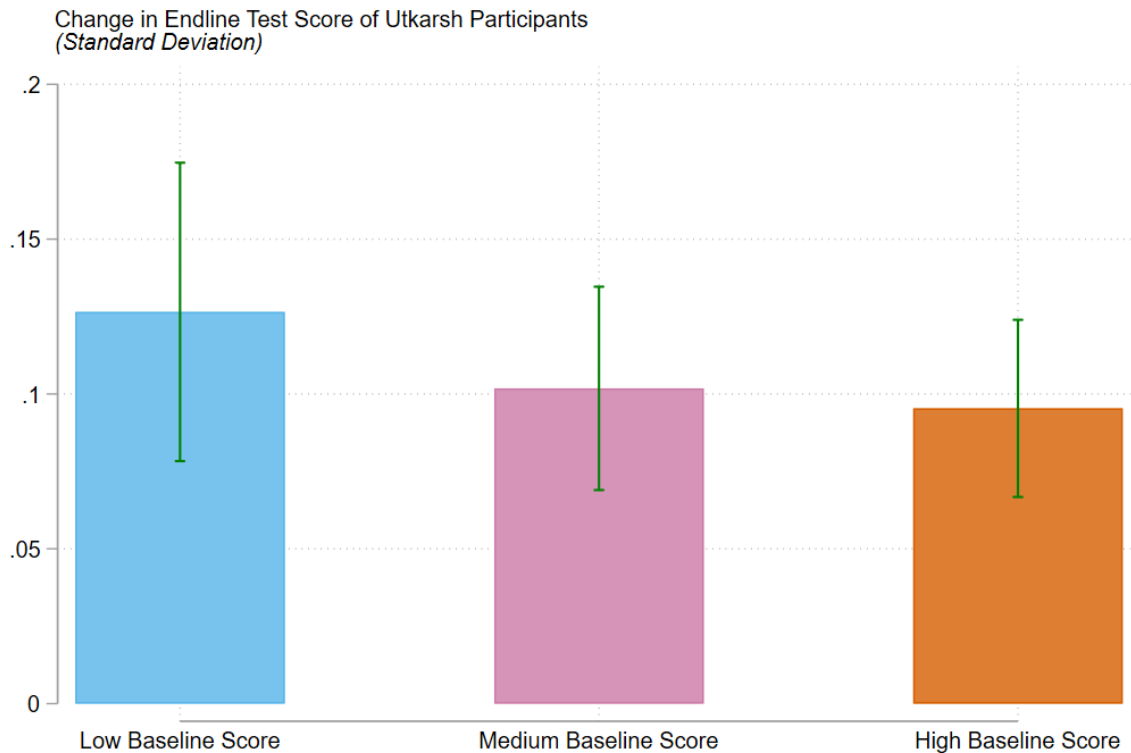
Figure 7: Treatment Effects by Baseline Test Score

One concern when measuring test score effects is that the intervention made students easier, or harder, to find at follow-up, leading to differential attrition by treatment status. Table 5 estimates the effect of Utkarsh on students being unable to be located for follow-up testing (i.e., attrition). In our setting, we find that students were about 2 percentage points less likely to be missing an endline test score in the Standard Utkarsh treatment. As we attempted multiple times to locate each student, this is not equivalent to a measurement of student attendance. Because of this differential attrition, we constructed treatment bounds following Lee (2009). In all cases both the magnitude and statistical significance are similar to the primary effects presented above.

Table 5: Attrition

	Missing Endline Score (1)
Standard Utkarsh	-0.023**
	(0.011)
Flexible Utkarsh	-0.013
	(0.011)
Observations	5,756
Control Mean	0.064
Flexible Utkarsh=Standard Utkarsh (p value)	0.266

Notes: Regression includes strata fixed effects, standardized IRT scores from baseline English, Math, and Odia tests, a dummy for being female, age of the pupil, dummies for missing control variables. Heteroskedasticity-robust standard errors, clustered by school, in parentheses. *** p<0.01 ** p<0.05 * p<0.1

In Table 6 we present additional student-level outcomes. In column 1 we tested for whether the intervention affected student attendance. Students were equally likely to attend school across all treatment arms, with about 66 percent of students present at the start of spot-check visits.

Students did not change their likelihood of taking private tuition (column 2) or the number of hours per week spent on private tuition (column 3). Students in the Standard Utkarsh schools did spend Rs27 per month less on private tuitions (column 4), statistically significant at the 10 percent level. This is an 8 percent decrease relative to the control group.

Almost all students across all three arms (95 percent) planned to take private tuition for their board exams and the interventions did not change these plans (column 5). Finally, students in the Standard Utkarsh treatment were 3 percentage points more likely to aspire to a Bachelor's degree or higher level of education (column 6).

Table 6: Additional Student-Level Outcomes

	Present at Spot Check	Taking Private Tuition	Time Spent on Private Tuition (Hour/week)	Money Spent on Private Tuition (Rs/month)	Planning to Take Private Tuition for Board Exam	Highest Level of Education Hope to Achieve: Bachelor's degree or above
	(1)	(2)	(3)	(4)	(5)	(6)
Standard Utkarsh	0.014	0.004	0.193	-26.501*	0.003	0.033**
	(0.022)	(0.012)	(0.386)	(13.911)	(0.007)	(0.016)
Flexible Utkarsh	0.012	0.002	0.069	-7.447	0.005	0.000
	(0.022)	(0.012)	(0.385)	(14.338)	(0.007)	(0.016)
Observations	5,710	5,445	5,444	5,388	5,431	4,767
Control mean	0.662	0.752	11.366	348.013	0.954	0.577
Flexible Utkarsh=Standard Utkarsh (p value)	0.936	0.846	0.747	0.152	0.721	0.036

Notes: All regressions include strata, week, and day-of-week fixed effects; standardized IRT weights from baseline English, Math, and Odia tests, age of the pupil, and indicator variables for endline interview phase. In addition, Column 1 includes spot-check visit phase fixed effects, Columns 2-5 include an indicator variable for taking private tuition in the baseline, Column 5 includes an indicator variable for planning to take private tuition for board exam in the baseline, and Column 6 includes an indicator variable for baseline level aspiration for highest level of education. Heteroskedasticity-robust standard errors, clustered by school, in parentheses. *** p<0.01 ** p<0.05 * p<0.1

5.2 Teachers and Classrooms

Based on data from the monitoring visits, nearly all teachers implemented the program. Almost all teachers in the both the Standard and Flexible Utkarsh treatments had used an Utkarsh worksheet on the most recent teaching day (98 and 94 percent). Part of the Flexible implementation required that teachers additionally fill out a flexible Utkarsh teaching plan—46 percent of teachers had done this. Over 70 percent of teachers in both treatment arms also conducted the correct Utkarsh lesson for the week.

In Table 7 we present other outcomes for teachers and classrooms. Teachers were equally likely across all three arms to be present in the classroom, 84 percent (column 1).

	Teacher was present in the classroom	Teaching and learning material visible in the classroom	At least one student had chance to express their own idea	Teacher asked question	Teacher answered questions properly	|-----Indices-----| Teacher familiar with content, encouraging students, and responding to student questions satisfactorily	Students behaved well	Quality of teaching practices
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Standard Utkarsh	0.017	0.014	0.214***	0.254	-0.082	0.278*	0.513***	0.713***
	(0.053)	(0.014)	(0.077)	(0.162)	(0.164)	(0.166)	(0.158)	(0.143)
Flexible Utkarsh	-0.055	0.013	0.138*	0.230	0.027	0.255	0.402**	0.715***
	(0.055)	(0.013)	(0.082)	(0.144)	(0.173)	(0.174)	(0.163)	(0.150)
Observations	299	249	249	249	249	249	249	249
Control mean	0.838	0.988	0.470	-0.116	-0.004	-0.199	-0.290	-0.484
Flexible Utkarsh=Stan-dard Utkarsh (p value)	0.181	0.809	0.341	0.874	0.384	0.894	0.447	0.989

Notes: Regression includes strata, week-of-survey, and day-of-week fixed effects, baseline classroom size, 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, participating in the Utthan program, and indicator variables for monitoring visit phase. All estimates in columns 2-8 are conditional on teacher being present in the classroom. Standard errors, clustered at school level, are in parentheses. Heteroskedasticity-robust standard errors, clustered by school, in parentheses. *** $p<0.01$ ** $p<0.05$ * $p<0.1$

Classrooms were also equally likely to have teaching and learning materials visible in the classroom (column 2). Note that almost all classrooms (99 percent) in the control group already had teaching and learning material visible. In both treatment arms at least one student was 21 percentage points (Standard) and 14 percentage points (Flexible) more likely to express their own idea (column 3), consistent with the program objective of creating a more engaging classroom.

We further created five indices to capture additional classroom characteristics. Based on these indices, the interventions did not change the Teacher Asked Questions or Teacher Answered Questions Properly indices (columns 4 and 5). Standard Utkarsh increased the value of the index that incorporates whether teachers were familiar with the content, encouraging students, and responding to student questions satisfactorily (column 6). Both interventions increased the Students Behaved Well and Overall Quality of Teaching Practices indices (column 7 and 8).

In the next two tables we present Students' and Teachers' Perceptions of Utkarsh. Table 8 shows that almost all student reported participating in Utkarsh. Of those who participated, about 90 percent reported enjoying it, 17 percent reported boredom, and about 72 percent thought that it would help with their board exams. Despite the high percentage who liked it and found it useful, about 68 percent reported that they preferred the regular curriculum to Utkarsh. We find no statistical difference across the two treatment arms (column 3).

Table 8: Student Perception of Utkarsh (only for Utkarsh schools)

	Observations	(1) Standard Utkarsh Mean/SE	(2) Flexible Utkarsh Mean/SE	(3) t-statistic of Difference (1)-(2)
Participated in Utkarsh	3625	0.993 (0.002)	0.985 (0.011)	0.009
Enjoyed Utkarsh	3584	0.897 (0.011)	0.912 (0.009)	-0.015
Got bored from Utkarsh	3581	0.173 (0.009)	0.171 (0.009)	0.002
Utkarsh will help in board exam	3584	0.739 (0.015)	0.717 (0.015)	0.023
Prefer regular curriculum to Utkarsh	3585	0.683 (0.014)	0.681 (0.013)	0.002

The value displayed for t-tests are differences in the means across the groups. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level

Teachers' perceptions of Utkarsh were mostly similar across the two versions of the program (Table 9). While about 72% of the teachers felt that the Utkarsh program was beneficial for the students, about 74% of the teachers felt pressure from parents to finish the regular curriculum. This indicates that communications with parents about the benefits of the program might be useful. Similarly, it may be useful to also communicate to teachers regarding the program benefits. The one statistically-significant difference in perceptions was for whether teachers thought the program was beneficial for themselves. This was 4.5 percentage points higher in the Flexible arm. That may suggest that teachers preferred the additional flexibility, although the rate of stating that the program is beneficial is already high in the Standard version (86%).

Table 9: Teacher Perception of Utkarsh (only for Utkarsh schools)

	Observations	(1) Standard Utkarsh Mean/SE	(2) Flexible Utkarsh Mean/SE	t-test Difference (1)-(2)
Felt pressure from parents to finish regular curriculum	569	0.745	0.717	0.027
		(0.028)	(0.030)	
Students benefitted from Utkarsh	572	0.752	0.724	0.028
		(0.031)	(0.029)	
Teachers benefitted from Utkarsh	571	0.860	0.905	-0.045*
		(0.021)	(0.017)	

The value displayed for t-tests are differences in the means across the groups. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level

5.3 Headmasters

When the enumerators arrived at each school, they noted whether the school's headmaster was present. Table 10 estimates whether the intervention changed this likelihood. We find that only 14 percent of headmasters were present in control-group schools. This rate was higher in both of the treatment schools, but the differences were not statistically significant.[12]

---

[12] Headmaster absence includes schools with a permanent headmaster who was absent and not having a permanent headmaster. The lack of an effect is similar if we condition on the school having an assigned headmaster.

Table 10: Headmaster Attendance

	Headmaster Present
	(1)
Standard Utkarsh	0.083
	(0.055)
Flexible Utkarsh	0.066
	(0.052)
Observations	298
Control mean	0.143
Flexible Utkarsh=Standard Utkarsh (p value)	0.759

Notes: All regressions include strata, monitoring visit phase, week-of-survey, and day-of-week fixed effects; a dummy for being female, age of the headmaster and age squared, years and experience and experience squared, a dummy for the school being multi-section, and total enrollment. Heteroskedasticity-robust standard errors, clustered by school, in parentheses. *** $p<0.01$ ** $p<0.05$ * $p<0.1$

Section 7.   CONCLUSION

When students fall behind in school they are often unable to catch up. In Odisha, class 9—the penultimate year of lower secondary school—is a critical year for students. People for Action designed the Utkarsh program to enable students who were behind grade level in class 9 to catch up to curriculum-level material. The goal is to prepare them for entering class 10, when students take board exams that determine whether they can continue on to upper secondary school.

Based on a randomized controlled trial in Odisha State, India, we found that Utkarsh increased student test scores on a test that included elements of English, Math, Odia, and Science as well as increasing each test score individually.

The Utkarsh model was implemented in two different, but related ways. In some randomly selected schools, teachers received the exact calendar of when to implement different lessons—this was the Standard Utkarsh model. In other randomly selected schools, teachers had the flexibility to implement the part of the program that they believed to be most salient to their students.

For student achievement, the average test score gains were about 0.1 SD, representing between 40 percent (Odia) and 200 percent (math) of what the control group learned over the same period. The test score gains between the two versions of Utkarsh were statistically indistinguishable.

The program improved student learning for all subjects. In most subjects the program benefitted students throughout the learning distribution, although in English students who were initially at lower competencies at baseline improved more. For the Science test, score gains were higher for the middle and top terciles, but we cannot reject that all the gains are equal.

Teachers in both treatment arms implemented the program and demonstrated equal knowledge about the program components. Teacher quality also increased.

Overall, the program achieved its goals of increasing student test scores across the entire distribution of initial student performance. We find no effect on student attendance of the program. While we find no effect on the likelihood of taking private tuitions, students in the Standard Utkarsh program reported spending 8 percent (Rs 27) less on private tuitions. Students in the Standard Utkarsh program also report slightly higher aspirations of completing a bachelor's degree or higher.

While the program did not change the likelihood of teachers being present in the classroom, we find that Utkarsh did change classroom operations, making them more engaging. We find that students are more likely to independently express an idea in Utkarsh classrooms, and that teachers in the Standard Utkarsh classroom are likely to answer questions properly and engage students. The quality of teaching practice also increased in all Utkarsh classrooms.

Section 8.   REFERENCES

Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukerji, S., Shotland, M., & Walton, M. (2017).

From Proof of Concept to Scalable Policies: Challenges and Solutions, with an Application. *Journal*

of Economic Perspectives, *31*(4), 73–102. https://doi.org/10.1257/jep.31.4.73

Beg, S., Fitzpatrick, A., Lucas, A., & Tsinigo, E. (2020). *Strengthening Teacher Accountability to Reach*

All Students (STARS)—Milestone 4: Impact Analysis Report.

http://documents.worldbank.org/curated/en/184851585721598666/pdf/Strengthening-Teacher-

Accountability-to-Reach-All-Students-STARS-Milestone-4-Impact-Analysis-Report.pdf

Evans, D. K., & Yuan, F. (2019). *Equivalent Years of Schooling: A Metric to Communicate*

Learning Gains in Concrete Terms (Working Paper No. 8752; Policy Research Working

Paper, p. 52). World Bank.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects.

The Review of Economic Studies, 76(3), 1071–1102.

TRANSFORM SCHOOLS

Head Office
CBR-II, First Floor, Khasra Number- 360,
Sultanpur, New Delhi - 110030, India.
partnerships@peopleforaction.org
www.transformschools.in

Transform Schools is registered in India as People For Action and
in the UK as Transform Schools UK