



Utkarsh (Secondary School Readiness Programme) – JPAL Evaluation

Pilot Report

March 2019

Executive Summary

Despite substantial gains in secondary school enrolment in India, the median Class 9 student does not have grade-level numeracy and English skills. Previous studies of primary school students have shown that targeting instruction at the level of student learning instead of the curriculum can improve learning outcomes. However, whether targeted instruction led by existing teachers a) improves learning outcomes for remedial secondary school learners, or b) helps students pass the high-stakes exams required to complete secondary school are open and important questions.

Utkarsh is a government-led remedial education programme for students in Class 9 in Odisha, India. This programme was designed by the Kusuma Foundation (now People for Action).¹ It groups students into “camps” according to their learning level where they work on level-appropriate skills, giving them the strong foundation that they need to tackle grade-level curriculum. This programme will be evaluated using a 300-school, multi-armed randomised controlled trial (RCT) and will provide lessons for a subsequent state-wide scale up of the programme.

To inform the design and implementation of the evaluation, a pilot was conducted in the 2018-2019 academic year. As part of evaluation activities, Abdul Latif Jameel Poverty Action Lab (J-PAL) enumeration teams visited 24 schools, interviewing 24 headmasters and 72 Class 9 teachers and interviewing and testing 622 Class 9 pupils in English, mathematics, Odia, and science. This report summarises the findings from the pilot. A comparison of scores between the treatment and control groups suggests that Utkarsh may increase learning. Mathematics, English and science test scores in the Utkarsh treatment group are approximately 0.27-0.29 standard deviations higher than in the control group, while Odia scores are 0.36 standard deviations higher. The average effect is about 0.3 standard deviations (SD) of test scores, which is quite large. According to Evans and Yuan (2019), a 0.3-SD gain is equivalent to between 1.7 and 2.5 years of typical progress in school. Due to limited statistical power, among other concerns, we caution that these are preliminary estimates and not statistically significant. The estimates, however, show real promise for a full evaluation. Nothing in the pilot data collection or analysis portends major issues for the research design or planned future data collection efforts for the full study.

¹ Utkarsh is the Odisha government's term for their implementation of Transform Learning, which was previously called the Secondary School Readiness Programme (SSRP). The SSRP was developed by the Kusuma Foundation with generous financial support from Kusuma Trust UK. KTUK's contribution has resulted in significant traction in an uncharted and important area of secondary education. A change in the Trust's strategy for India meant that it ceased its India operations from 31 December 2018. The Kusuma Foundation's core team has set up the broader Transform Initiative to take this work forward and impact the lives of millions of children via People for Action, an Indian NGO with startup support from KTUK.

Contents

Executive Summary	1
Content	2
List of Tables	3
List of Figures	3
Section 1: Introduction and Background of the Utkarsh Evaluation	4
1.1 Introduction	4
1.2 Background	4
1.3 The Utkarsh Model	4
1.4 Pilot Aims	5
Section 2: Pilot Evaluation Design	6
2.1 Theory of Change	6
2.2 Pilot Design	6
2.3 Pilot Scope	6
2.4 Sample and Sampling Procedure	7
2.5 Outcome Measures	7
2.6 Data Collection Instruments and Content	7
2.6.1 Field staff recruitment and training	8
2.6.2 Baseline coverage	8
2.6.3 Endline coverage	8
2.6.4 Data Collection Challenges	9
Section 3: Data Analysis Summaries	9
3.1 Description of Baseline Data	9
3.1.1 Headmasters	9
3.1.2 Teachers	10
3.1.3 Students	10
3.2 Baseline Equivalence on Time-Invariant Measures	10
3.3 Description of Monitoring Visit Data	10
3.4 Description of Endline Data	11
3.5 Description of Utkarsh Leveling Exam Data	13
3.6 Summaries from Focus Groups	14
Section 4: Lessons Learned	15
4.1 Key Lessons from the Pilot for Implementation	15
4.2 Achievement of Pilot Objectives for Proposed Full Study	15
Section 5: Conclusion and Next Steps	16
5.1 Conclusions	16
5.2 Next Steps	17
Appendix	18

List of Tables

Table 1: Baseline Survey Completion Rate by Instrument	8
Table 2: Endline Coverage Rate by Instrument	9
Appendix Table 1: Baseline Summary Statistics of Headmasters	18
Appendix Table 2: Baseline Summary Statistics of Teachers	19
Appendix Table 3: Baseline Summary Statistics of Students	20
Appendix Table 4: Monitoring Visit Summary Statistics of Teacher Survey	21
Appendix Table 5: Monitoring Visit Summary Statistics of Teacher Survey (Utkarsh only)	22
Appendix Table 6. Monitoring Visit Summary Statistics of Teacher Survey (Utkarsh only) cont.	23
Appendix Table 7. Effect of Utkarsh on Endline Test Scores	24

List of Figures

Figure 1: Pilot Study Design	6
Figure 2: Bar Charts of Test Scores by Study Arm	11
Figure 3: Histograms of Test Scores by Study Arm	12
Figure 4: Utkarsh Levelling Exam Scores vs. Pilot Baseline Exam Scores	13
Figure 5: Pilot Baseline Exam Scores vs. Utkarsh level Assignments	14

Section 1: Introduction and Background of the Utkarsh Evaluation

1.1 Introduction

Despite tremendous progress in expanding access to primary and secondary school, many secondary school students in India lack basic foundational knowledge. This lack of preparation inhibits mastery of the advanced material needed to complete secondary schooling or pass the high-stakes board exams at the end of Class 10. Recent results from programmes in primary schools show that targeting teaching to the level of the student instead of the official curriculum is a promising way to transform student learning (Banerjee et al., 2017; Duflo, Kiessel, and Lucas 2018). The Secondary School Readiness Programme (Utkarsh) applies this approach at the secondary school level, and has the potential to improve educational attainment, particularly for remedial learners. We plan to conduct a randomised controlled trial to evaluate the Utkarsh programme in Odisha State during the 2019-2020 academic year. To inform the design of this larger experiment we conducted a 24 school pilot study during the 2018-2019 academic year. Here we present the main outcomes and lessons learned from the pilot evaluation.

1.2 Background

Odisha State, the setting for this study, is one of the poorest states in India and is in the bottom quartile of literacy rates (Reserve Bank of India 2017 and 2018). Gross secondary enrolment is 77%, below the national average (National University of Educational Planning and Administration, 2016). Odisha has a substantial population of minority and tribal groups, approximately 40% of enrolled students, making our study particularly relevant for settings with other marginalised groups.

In Odisha, students start school at age 6. Primary school covers Classes 1-5 and middle school covers Classes 6-8. Classes 9-10 are lower secondary school while Classes 11-12 are upper secondary school. Class 10 is a crucial transition point for Indian students. Board examinations, which determine entrance into higher secondary grades, occur at the end of Class 10. Students, especially girls, often drop out after completing lower secondary grades. Based on the 2013-14 graduating cohort, only 50% of students who started lower secondary school completed higher secondary school in Odisha. Some of this excessive drop-out is due to poor schooling quality, particularly for students at the lower end of the learning distribution. Strict adherence to the official level of the curriculum causes some students to fall further behind, with learning gaps growing each year. Therefore, while grade-specific learning standards are important, teaching to children at their own level to decrease the gap between their own knowledge and the official standards is also crucial and offers an inclusive approach to learning.

Student achievement in Odisha is both low and heterogeneous. Das and Zajonc (2010) find 50% of the enrolled students in Odisha fail to meet a basic international benchmark of mathematical knowledge. Based on Utkarsh monitoring data from Hardoi district, no Class 9 students achieved even a Class 8 competency level. In stark contrast, the top 5% of students in Odisha outperform the same group from other low-income countries. Improving the quality of teaching and providing additional support to students behind grade level can bridge these learning gaps and raise overall achievement. Implementing these interventions in Class 9, shortly before students arrive at the crucial transition to upper secondary school, could improve their performance on high-stakes standardised tests, producing far-reaching effects on later grade progression and labor market outcomes.

1.3 The Utkarsh Model

The Utkarsh programme has both teacher and classroom components. At the classroom level, content is provided to offer remedial skills in the first part of the programme, successively progressing to grade level content. Existing teachers receive training, materials, and support to implement the programme in their Class 9 classrooms.

From the outset, the programme uses existing government employees. The Transform Trust trains district-level teams of Master Trainers. Master Trainers then provide training and support to existing Class 9 teachers and headmasters from programme 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. Teachers and students receive these handbooks for each subject. The Transform Trust provides technical and financial assistance for the programme and partners with the government to maintain quality.

Classroom teachers return to their schools and implement the programme. First, to establish existing learning levels, teachers conduct grade levelling tests for all their students. The exams cover Odia (the local language), English, science, and mathematics. Second, teachers assign students an effective grade level based on test scores and teacher input. Finally, students engage with the Utkarsh programme at the relevant time for their learning level.

The lowest level learners, those who test at the Class 5 level or below, participate in Foundation Camp near the start of the school year. In the Foundation Camp, students receive four hours of remedial instruction every day over 18 instructional days. Students who perform above the Class 5 level participate in alternative self-learning activities during the remedial classes.²

Once students complete the Foundation Camp, they are combined with students assessed to be at learning levels between Class 6 and 8 for the Supported Learning Phase. During this phase, students receive three hours of remedial teaching each day for 45 instructional days. Students at grade level continue with self-learning activities.

The final phase of the programme is the Consolidation Camp, which is intended to prepare all students for the transition to Class 10 and the Class 10 board examination. This phase involves 18 hours of Class 9 material over six days. At the end of the Consolidation Camp phase, teachers again assess all students on their learning through levelling tests.

The pilot study provided valuable insights for designing the broader rollout and evaluation of the programme. In our pilot phase, we evaluated the feasibility of conducting and evaluating the Utkarsh model in government schools.

1.4 Pilot Aims

Our pilot aimed to answer the following questions on both the programme and its evaluation:

Aim A1: what is the most effective and cost-effective way to conduct a baseline and endline, including student assessments? Because of a lack of administrative data in Odisha on Class 9 students, high estimated student attrition, and an interest in examining treatment effect heterogeneity by baseline test scores and basic demographics, efficiently conducting both a baseline and endline is crucial for the larger study. We tested two different models of assessing student achievement across both treatment and control schools: a) administering group exams and b) administering a one-on-one exam. Further, we tested the administering of surveys to school headmasters, teachers, and students as well as a classroom observation tool.

Aim A2: based on the existing levelling exam, are students placed in the correct levels? Can levelling exams be shortened? Teachers both conduct and grade the exams in the treatment schools. We collected these exams and audited the grading and student placement during monitoring visits. This allowed us to compare outcomes from the levelling exams to independently designed student assessments to validate the different forms of assessment. Furthermore, focus group discussions (see below) were conducted to gauge whether the multi-hour length of the levelling exams is a barrier to implementation.

Aim A3: how can we effectively measure whether teachers in treatment schools implement the Utkarsh programme, i.e. programme take-up and fidelity of implementation? Both Duflo, Kiessel, and Lucas (2018) and Banerjee et al. (2017) highlight potential pitfalls when targeted instruction programmes are implemented without extra monitoring and support. We conducted one monitoring visit during the programme phase to measure adherence to the model. We conducted two focus groups with students and one focus group with teachers who have participated in the previous year of Utkarsh to understand their experiences, the challenges that the programme faces, and identify roadblocks to evaluation.

Aim A4: can we follow-up with the same students from baseline? A primary challenge in completing an evaluation is tracking the same respondents over time, particularly in settings with high absenteeism. We tested protocols to successfully track students over time. Given the levels of student attrition we will determine whether tracking students outside of school is necessary. High student attrition is a specific concern for the Transform Trust and the research funder (the Kusuma Trust UK). In this pilot, we attempted to follow the same students at each visit, but did not attempt to track students outside of school.

Aim A5: what is the correct sample size for a full-scale RCT and what is the budget necessary to ensure complete data collection? Based on data collected in our achievement exams and spot checks we calculated estimates of the minimum detectable effects for different sample sizes. By developing protocols and estimates for survey team organisation, exam grading and data entry, we can more precisely estimate costs of a full evaluation.

Aim A6: what are the barriers, if any, to successful evaluation? What other aspects of the programme and local context could pose a challenge to ensuring the success of the large-scale randomised trial? By conducting surveys, collecting data from respondents, and understanding the details of programme implementation we may identify issues that would arise during the full study.

Section 2: Pilot Evaluation Design

2.1 Theory of Change

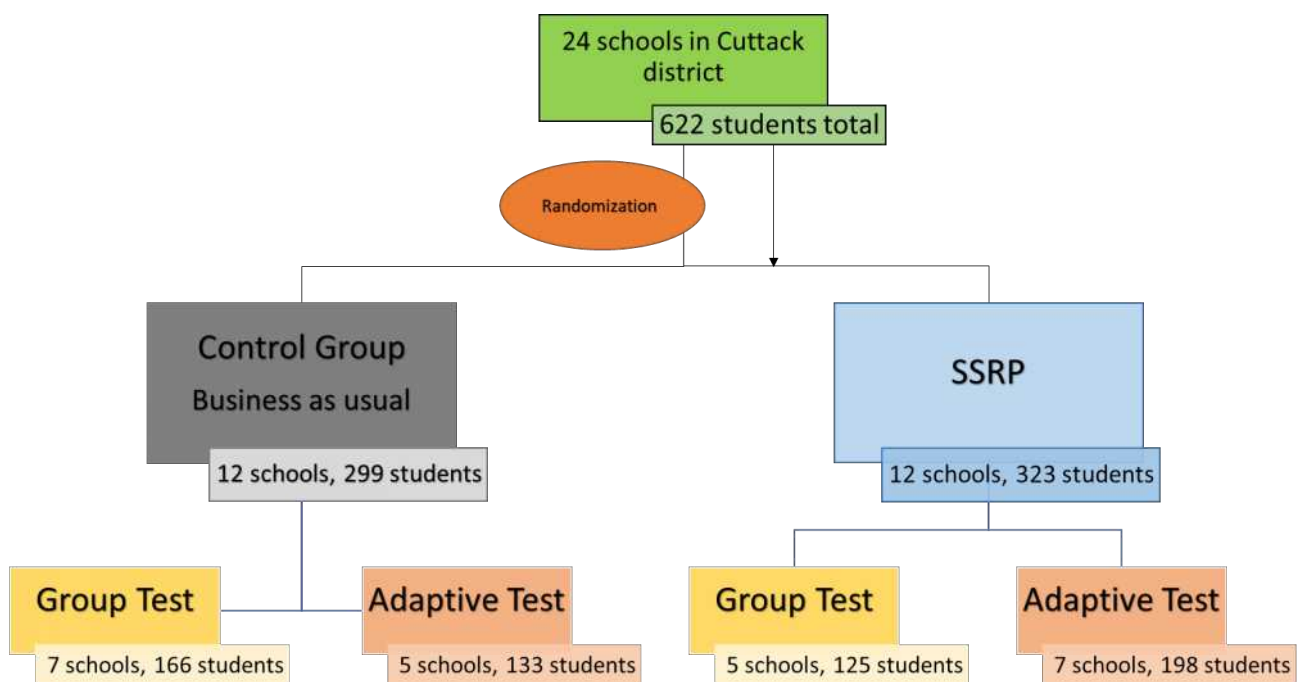
Students are increasingly in school, but not learning, especially those from marginalised groups. Moreover, students are at different levels of preparation, particularly first-generation learners. Therefore, they do not benefit from instruction that is beyond their existing learning level. As a result of this heterogeneity within the classroom, teachers are unable to effectively teach grade-level material to all students

The Utkarsh intervention aims to increase learning levels for all students, and will ensure learners of all genders and backgrounds have the same opportunities. Utkarsh offers camps to Class 9 students that teach skills for different learning levels, building from remedial skills to grade level instruction. Our theory of change is that teaching material geared to the learning level of students reduces within-class heterogeneity and allows teachers to effectively build grade-level skills. During Utkarsh hours, students who are behind receive instruction at their level, remedying basic skills and allowing them to catch-up to grade-level work. Teachers receive training in this method of teaching and a suite of teaching and learning materials, allowing them to focus on improving student learning given their existing skills.

2.2 Pilot Design

We piloted the Utkarsh tools and procedures in 24 schools in Cuttack district. The pilot was randomised to include 12 treatment schools and 12 control schools. The study design is pictured in Figure 1.

Figure 1 Pilot Study Design



2.3 Pilot Scope

To analyse the effects of the programme, we conducted three rounds of data collection – a baseline, a monitoring visit, and an endline. We completed the baseline survey in November 2018.

We also conducted monitoring visits of classroom behavior in 12 schools in January 2019, and a follow-up study in February 2019.

2.4 Sample and Sampling Procedure

We drew our sample from 280 secondary schools in Cuttack district, within 20 km of the district center.³ We randomly selected 24 schools from this list of schools after stratifying on total enrolment. This ensured that the final sample consisted of schools of varying sizes.

Within each randomly-sampled school, we drew the sample of survey respondents and exam participants as follows:

- a. Headmasters from each sampled school (12 treatment and 12 control for a total of 24) were automatically enlisted into the Utkarsh evaluation.
- b. Two Class 9 teachers were interviewed from each control and treatment school. The teachers were selected from the list of Class 9 teachers previously provided to us by the Transform Trust. We selected mathematics and English teachers first. If more than one of either type of teacher was present at the school, we chose randomly. In cases where two mathematics or English teachers were not present, a back-up teacher was selected randomly and interviewed.
- c. Thirty students who were present on the day of the visit and listed on the Class 9 register were randomly sampled from each school. These students were tested and interviewed. If a particular class had more than 30 students, only 30 were randomly sampled. If less than 30 students were present in Class 9, all students were automatically included. The total number of students enlisted into the Utkarsh pilot was 622 students. If a school was multi-section, we randomly selected one section and only randomly selected students within one section.

Half of the selected schools were randomly assigned to receive Utkarsh (treatment group); the other half were the control group. Then, within both the Utkarsh and the control groups, half of schools received group tests and half of schools received one-on-one tests.

2.5 Outcome Measures

Our primary outcome measures are:

- Did schools implement Utkarsh?
- Did student learning improve, as measured by performance on the mathematics, English, Odia, and Science achievement tests?
- Did teacher, headmaster, or pupil attendance improve?

2.6 Data Collection Instruments and Content

Our data collection involved surveys of three respondent types: 1) headmasters, (2) teachers, and (3) students. Notably, we used seven survey instruments for the baseline and endline surveys. These were the Headmaster Survey, Teacher Survey, Student Survey, and Student Assessment for mathematics, English, Odia, and Science. Additionally, we administered three instruments during the monitoring visit: a classroom observation tool and surveys of teachers and headmasters. The Principal Investigators (PIs) designed all the tools. The draft surveys went through a series of iterations before they were finalised. This involved adapting survey question to the Indian context and pre-testing the surveys to ensure questions were clearly worded 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 finalising them.

All surveys were administered in Odia. The surveys for the baseline were administered on paper. For the endline surveys and the monitoring visit, survey responses were directly entered into tablets. All programmed surveys were bench-tested to ensure that all consistency checks and skip patterns were working as expected.

The following presents the contents of each of the baseline and endline surveys:

- a. *Headmaster Survey:* The Headmaster Survey was administered to the headmasters in the sampled schools through an in-person interview. The Headmaster Survey sought information on the demographic characteristics, previous education/experience, daily activities, perceptions about teaching and learning in schools, and work-related stress and burnout.
- b. *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, and work-

c. *Student Survey*: The Student Survey was administered to selected Class 9 students. The questionnaire captured information on the students' background characteristics and aspirations and students' feedback on classroom teaching.

d. *Student Assessment for mathematics, English, Odia, and science*: The Student Assessments were administered to selected Class 9 students. These instruments were adapted from a combination of resources, including the Utkarsh assessment tools and handbooks and other standardised exams for secondary students.

2.6.1 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 field supervisors, team leaders, auditors, and interviewers/student assessors. A total of 49 field staff qualified for the fieldwork and were subsequently selected.

Different training sessions relating to pilot survey training, supervisor training, and general training were conducted. Training included both classroom and field practice. Classroom lessons were presentations, questions and answers, group discussions, and role-plays. Pilot baseline training was held from 4th to 9th November 2018 (i.e., 6 days of training) in a rented venue in Bhubaneswar. Forty-four people were trained on the baseline surveys and assessments. Only 36 surveyors were retained for the actual pilot study. A one-day supervisor training was held for 4 pre-selected candidates for the position of field supervisor. This was to help them have a general overview of the entire baseline survey and enhance their support of enumerators. This training took place on November 10, 2018, a day after the general training. The survey field team comprised 40 field staff members. The entire field team consisted of 4 survey field teams with each team composed of one supervisor and nine surveyors.

2.6.2 Baseline Coverage

Table 1 presents the baseline coverage rates for the instruments administered to the Headmasters, teachers, and students. We had an average coverage rate of 90.2% for all the data collection instruments (Headmaster Survey, Teacher Survey, Student Survey, mathematics, English, Odia, and science assessments). Completion rates were less than target rates because some schools lacked 30 Class 9 children; some students refused; some students were special needs children.

Table 1: Baseline Survey Completion Rates by Instrument

Instruments	Target	Completed	Refused
Headmaster Survey	24	24	0
Teacher Survey	48	48	0
Student Survey	720	622	0
Pupil Assessment [English]	720	622	0
Pupil Assessment [Math]	720	622	0
Pupil Assessment [Science]	720	621	1
Pupil Assessment [Odia]	720	621	1

2.6.3 Endline Coverage

Table 2 presents the endline coverage rates for the instruments administered to the headmasters, teachers, and students. While generally sample sizes are comparable for adult respondents (headmasters and teachers), we were only able to interview approximately 71 percent of target children. Note that baseline teachers who were absent or late the day of the endline were replaced, so that we would have sufficient sample size for that survey. Children lost to endline were absent the day of the survey activities; we discuss attendance and study attrition in a later section.

Table 2: Endline Survey Coverage Rates by Instrument

Instruments	Target	Completed	Refused	% Coverage
Headmaster Survey	24	23	1	95.8%
Teacher Survey	48	48	0	100%
Student Survey	622	441	0	70.9%
Pupil Assessment [English]	622	441	0	70.9%
Pupil Assessment [Math]	622	441	0	70.9%
Pupil Assessment [Science]	621	439	0	70.7%
Pupil Assessment [Odia]	621	438	0	70.5%

2.6.4 Data collection challenges

The research team encountered the following challenges:

- a. In many schools, there were limited or no school resources to support evaluation activities. The missing resources included extra classrooms, chairs, and tables, which hindered the setting up of a conducive environment for the interviews or pupil assessment activities. In particular, in schools with limited seating for assessments there were reports of suspected cheating by enumerators. Cheating was observed to occur more frequently when more than one student sat at the same desk. For that reason in the full study we will take steps to reduce student cheating by (for example) limiting the sample size per school so that assessments and randomly assigning students to seats.
- b. In some schools, there was limited enrolment and high pupil absenteeism. Some schools had fewer than 15 children per section and hence the team could not meet the target of 30 children per school. In addition, there was high absenteeism at endline.
- c. Rural areas were difficult to find and navigate. Poor road networks made transportation to schools extremely difficult and expensive in some cases. In one instance, despite contacting the school ahead of time to arrange the visit, one study schools' address was interchanged with another. The schools had been scheduled to be visited on different days. In order to stick to the tight study timeline, the enumerators decided to conduct the days' activities as planned even though one school was assigned to a group test and the other to a one-on-one test. For that reason, there are 13 group and 11 one-on-one schools in the sample.

Section 3: Data Analysis Summaries

3.1 Description of Baseline Data

Overall, the data that we collected appears to be high quality. In this subsection, we describe the means across the various measures and respondents. The next subsection addresses the similarities between the various groups by treatment status, i.e. baseline balance.

3.1.1 Headmasters⁴

We interviewed 24 headmasters, 14 males and 10 females. The average age of the headmasters is 51.4 years. Over 95 percent of them have a bachelor's degree or more. On average a headmaster worked for 19.3 years as a subject teacher before becoming the headmaster. The median length of their tenure as headmaster is 4 years, while the average is 7.6 years. We also asked the headmasters about the number of teachers employed in their schools. The averageschool had 11 teachers, about 7 of whom were Class 9 teachers.

⁴ See Appendix: Table I for more details on the data collected during the baseline, including these averages disaggregated by treatment group.

3.1.2 Teachers⁵

We interviewed a total of 48 Class 9 teachers – two teachers from each school. About 55 percent of the teachers are female, and on average the teachers are 39.4 years old. Ninety-five percent of the teachers have at least a bachelor degree. Teachers in our sample have on average worked for 13.5 years, 9.5 of which were in their current place of employment. We also asked teachers about trainings they have attended. Teachers in Utkarsh schools have attended more trainings (1.3 on average) than teachers in non-Utkarsh schools (0.7 on average).

3.1.3 Students⁶

Students in our study completed tests in mathematics, English, Odia, and Science. We used two types of tests:

1) a partially adaptive test where students all took the same initial questions, then based on their answers were then given either an additional set of harder or easier questions;

2) a group test where all students took the entire set of questions. We used item response theory (IRT) to solve for their latent score and put all students on the same scale. We then standardised this score to mean 0 and standard deviation 1 within each subject. Therefore, by design, the average score of this test should be close to 0. When presenting histograms of the test results, we shift the distribution to a mean of 100 and a standard deviation of 15, for simplicity of presentation.

In terms of demographic characteristics, just under half of our students are female (47%) and they are on average 12 years old. About 40 percent of students come from other backward castes, 30 percent students come from scheduled castes or scheduled tribes; the rest of the students come from general castes. Students report about one third of mothers are literate (37%) and just over half of their fathers are literate (54%). Almost all students want to attend senior high school (95%) and students report being absent about half a day in the previous week (0.6 days). 67% of students report liking their math class very much and 44% like English class very much.

3.2 Baseline Equivalence on Time-Invariant Measures

In addition to providing insight into means across the sample, we also tested for the equivalence of means across the treatment arms among time-invariant characteristics. The last column of tables 1, 2, and 3 reports the p-value of the t-test of the difference of mean between treatment arm and the control arm. If these p-values are above a certain threshold, most commonly 0.05, we can verify that the treatment group and control group are not systematically different, thus validating that randomisation was done properly.

As we were not able to field the baseline pilot prior to the intervention due to delays, some characteristics that may respond to the introduction of the programme may differ across treatment and control schools even at baseline. For instance, we asked students whether they had learned something interesting the day before the survey. The answer to this question could potentially be influenced by the introduction of the treatment, and therefore cannot validly be tested for balance across study arms. However, we can still test whether the randomisation procedure created comparable groups by looking at time-invariant characteristics (e.g. gender, language spoken at home), or time-variant measures that are unaffected by the treatment (e.g. age, length of travel from home to school). Looking the tables of baseline characteristics, we can note that none of the variables have a p-value of less than 0.05. This suggests that our randomization strategy was successful. This is encouraging for the full study: if we sample enough schools and students to guarantee statistical power, we would be able to make meaningful judgments about the effectiveness of the programme.

3.3 Description of Monitoring Visit Data

We conducted monitoring visits in three treatment and three control schools in January 2019. This number was chosen due to the limited time available and because this number allowed us to still achieve our objectives while fitting in our pilot budget. The goal of the monitoring visits was to measure compliance with the programme. Appendix Tables 4, 5 and 6 report the summary of our findings during monitoring visits.

⁵ See Appendix: Table 2 for more details on the data collected during the baseline.

⁶ See Appendix: Table 3 for more details on the data collected during Baseline.

Outcome 1: The majority of schools were implementing Utkarsh

We find that most of the Utkarsh teachers received training, while all of them received the Utkarsh manual. The schools were on track to complete the Utkarsh programme. All the Utkarsh schools completed the Foundation Camp phase, and they were conducting the Supported Learning Phase during the monitoring visits as scheduled. While the teachers understood who should attend the Foundation Camp, they were fairly confused about who should attend the Supported Learning Phase: two-thirds of teachers thought all students should attend that phase of the programme.

We found that all the Class 9 students of the school attended the Foundation Camp. Though a majority of the teachers reported that they follow the Utkarsh lessons, only half of the teachers were following the scheduled Utkarsh lesson for the day of our monitoring visit. Many of the teachers find Utkarsh lessons and activities to be difficult (50%).

We find that only half of the treatment school teachers were covering the official Class 9 curriculum, while all the teachers in the control school were following it. None of the control school teachers felt that they were behind schedule, while none of the treatment school teachers felt that they were ahead of schedule.

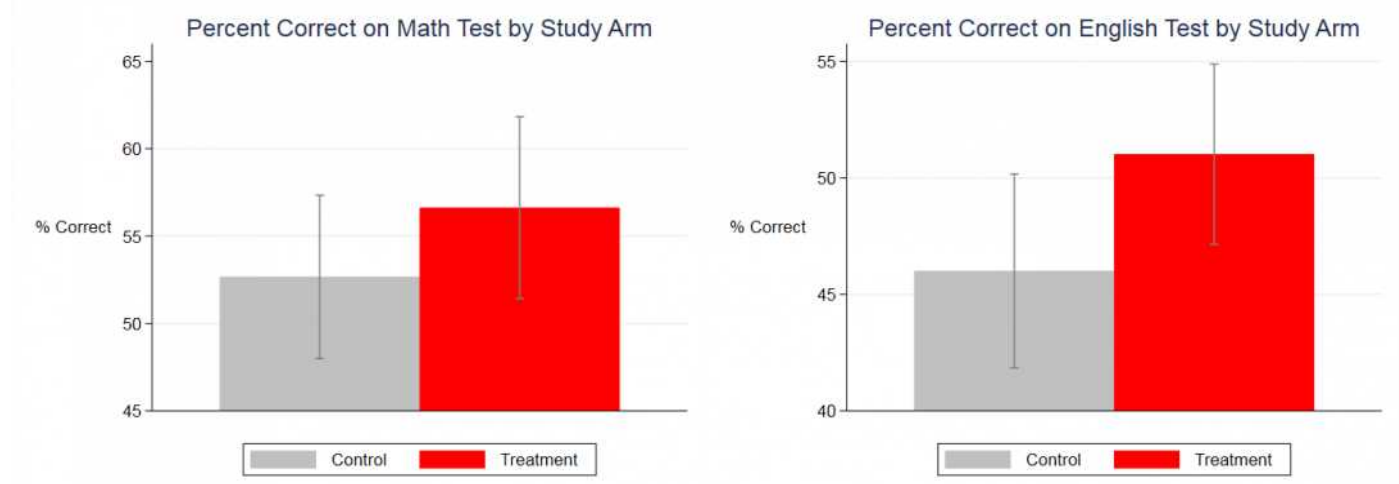
3.4 Description of Endline Data

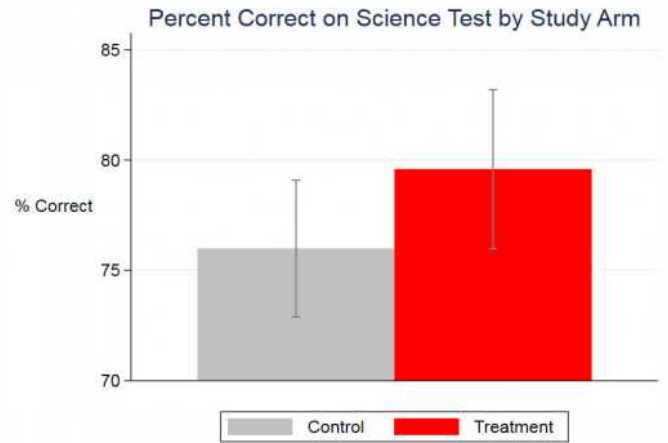
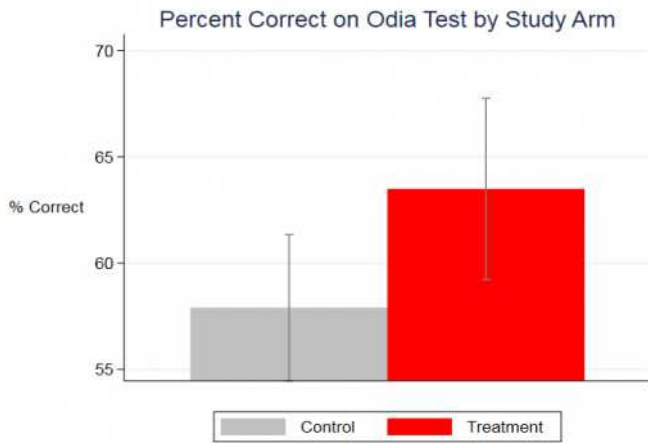
Outcome 2: Utkarsh may increase student test scores in all subjects

Our preliminary analysis of the endline data suggests that the Utkarsh intervention may improve learning. These results need to be interpreted with caution for four reasons. First, the small sample size for the pilot means we cannot rule out that these results would have happened by chance. Second, the shortened length of the intervention means that these effects may not be representative of how effective the full-length Utkarsh intervention would be. Third, we were only able to sample students for the pilot after the Utkarsh began, so our selection of students could have potentially been affected by the intervention. For example, if Utkarsh improves attendance in particular for high-performing students then our results are larger compared to what we would expect in the full study. If Utkarsh improves attendance in particular for low-performing students then our results are potentially underestimates of the true effect of the programme. Fourth, we plan to modify some of our current protocols for the full study to limit cheating, enumerator effects, and consistent grading in order to ensure a scientifically valid and rigorous evaluation.

With those potential issues noted, the intervention appears to raise test scores in all four subjects. Figure 2 presents the percent of questions students got right on each test for both the treatment and control groups. The whiskers on the bars show 95% confidence intervals; these are overlapping, indicating that the evident difference could have occurred by chance. Mathematics, English and Science test scores in the Utkarsh treatment group are approximately 0.27-0.29 standard deviations higher than in the control group, while Odia scores are 0.36 standard deviations higher (Appendix: Table 7). The average effect is about 0.3 standard deviations (SD) of test scores, which is quite large. According to Evans and Yuan (2019), a 0.3-SD gain is equivalent to between 1.7 and 2.5 years of typical progress in school.

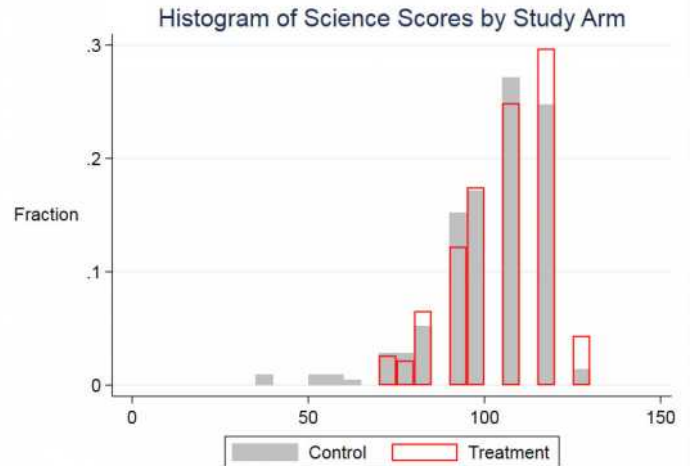
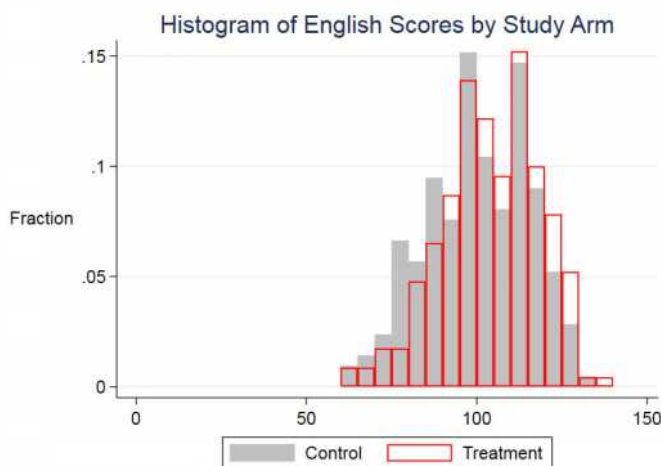
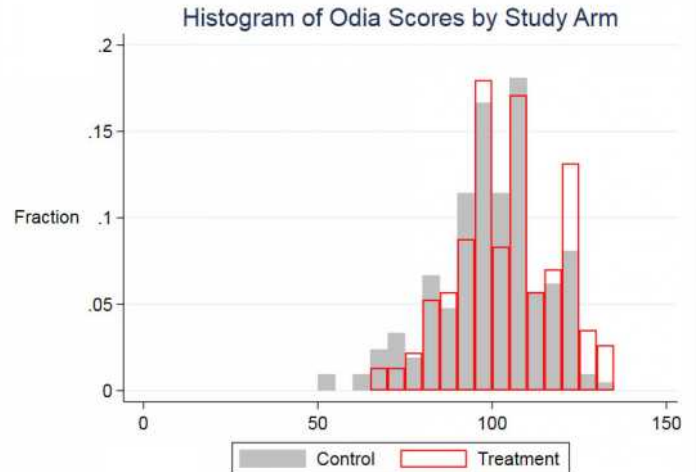
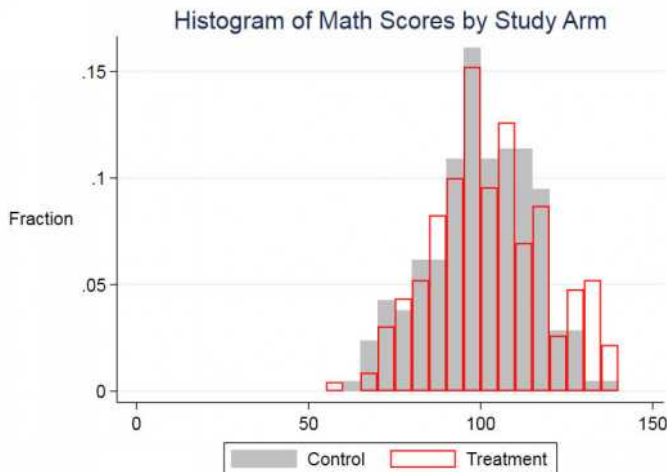
Figure 2: Bar Charts of Test Scores by Study Arm





Moreover, these gains occur throughout the distribution. Figure 3 shows histograms of test scores by study arm for each subject. Virtually all students improve their performance as a result of the intervention. This is fairly uncommon: most education programmes help just the best students or just the weakest ones. The design of Utkarsh likely contributes to these broad-based gains. By targeting instruction sequentially at the learning levels of different groups of students, it appears to be increasing performance across the board. If verified in a full-scale study, this finding would be a significant contribution to understand how to improve learning.

Figure 3: Histograms of Test Scores by Study Arm



Finally, our data indicate that several headmasters work in multiple schools due to vacancies. At the endline, 12 percent of headmasters report working in multiple schools. For the full study, we are concerned that headmasters may cross treatment status, and thus are working to reduced that likelihood in the full study.

Outcome 3: Utkarsh has little effect on student attendance

Enumerators reported that our procedures for matching students over time were sufficient. In addition to pictures we had validation questions (name; sibling name; etc.) that allow us to confidently state that our evaluation will be able to link students over time. However, our pilot identified that attrition will likely be a challenge in the full study. High rates of absenteeism and therefore attrition are common problems in education studies, and not unique to this evaluation or setting.

In Utkarsh schools, the attrition rate was 28.8 percent while in control schools the attrition rate was 29.4 percent. However, because our baseline sampled students who were present at the end of the Foundation Camp, we believe that this may be an underestimate of what we may expect in the full study. Absenteeism, meaning the percent of all enrolled students (not just in our study sample) who are absent on the day of the endline is high: in Utkarsh schools, 49.5 percent of students are absent while in control schools 54.6 percent of students are absent. It is encouraging from a programme perspective that Utkarsh may increase student attendance. However, that success may still pose challenges for evaluating the impact of the programme. While rates are not substantially different between the treatment and control groups in this pilot, in a larger study these differences could be exacerbated and determined to be statistically significant. We will take steps to reduce the likelihood of differential attrition in the large-scale evaluation by devoting additional resources for tracking students.

3.5 Description of Utkarsh Levelling Exam Data

To compare the Utkarsh levelling data to our data we collected, graded, and entered levelling exams from 35 students in 4 subjects across 4 schools. This exercise was labor-intensive and took 9 monitors 9 days (including training). Our comparison of the Utkarsh levelling exam data to the baseline exams from the pilot study reveals a promising pattern: it may be possible to accurately place students into ability levels using much-shorter exams. In particular, our group tests for English had a strong correlation with both the Utkarsh levelling exams for English, and also the assigned levels for English. Figure 4, below, shows how the pilot baseline exam score related to the raw test score from the levelling exam, while Figure 5 shows the relationship between the pilot baseline exam score and the student's assigned level. Statistical tests confirm this relationship: we can predict 81 percent of the variation in Utkarsh levelling test scores based on the student's pilot baseline score.

This finding has several caveats. First, it is based on just 35 students worth of data, because that is the number for whom we were able to match across exam types during the pilot. Second, the correspondence between the Utkarsh levelling tests and the pilot baseline scores for the other subjects was much weaker. Third, we refined the pilot tests substantially between the baseline and the endline, and thus the quality of the data we are able to get from them is likely to improve. This may allow us to better match the scores from the Utkarsh exams. If this result holds up in the full study, however, it could be very useful for scaling up the Utkarsh, as the lengthy exams are a major logistical hurdle for implementing the programme.

Our key takeaway from this exercise is that collecting and regrading the levelling exams for the full study is likely infeasible from a budgetary perspective, but that the overall evaluation could potentially incorporate the levelling exams from the programme if programme data collected additional identifiable information to link students with the evaluation. Such an exercise would need to be carefully considered and implemented to protect study children's data privacy.

Figure 4: Utkarsh Levelling Exam Scores vs. Pilot Baseline Exam

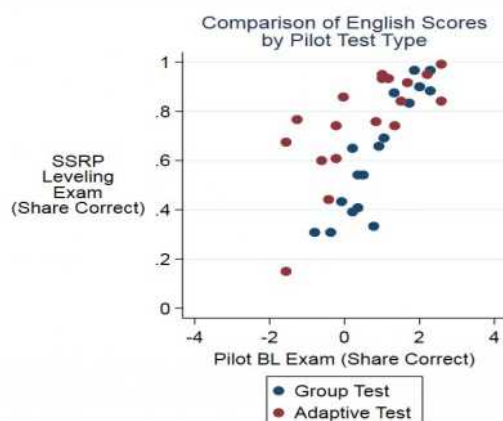
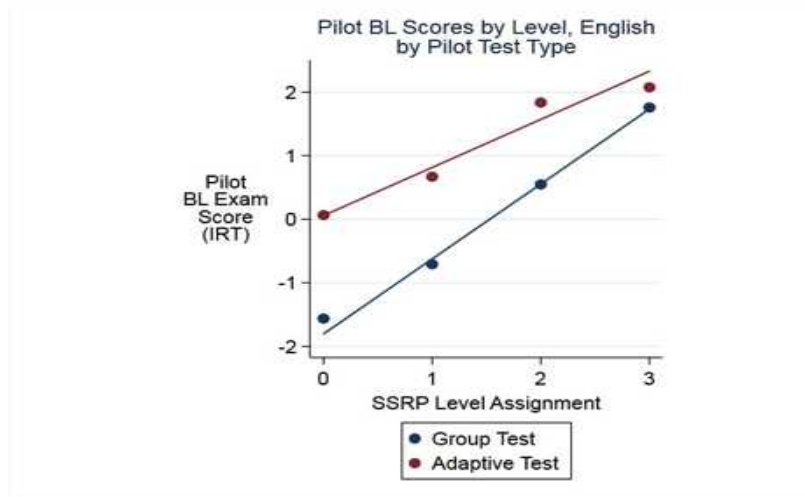


Figure 5: Pilot Baseline Exam Scores vs Utkarsh Level Assignments



The current Utkarsh implementation collects a large amount of data that could possibly be used for additional analysis. If the levelling exam cover sheet collected additional information on identifying variables then it would be easier to link the levelling data with the evaluation data for the full study. For example, if the child's date of birth, parent/sibling names or other variables that could be used as checks against name merges, we could potentially maximise the usefulness of administrative data.

3.6 Summaries from Focus Group

Focus group discussions with students and teachers were conducted in 2 schools in Sambalpur district with a total of 18 students and 12 of teachers. They covered a broad range of topics and ideas on schooling in general and the Utkarsh programme specifically.

Students consider the Class 10 exams to be crucial for job market opportunities and for pursuing further schooling. At least for the students participating in the camps, they are motivated to learn. There is some demand for additional instructional time for difficult and confusing topics to help students prepare for Class 10 exams. As also came through in the survey data, private tuitions are very common and also expensive: students spend 4-6 hours a week on attending them. In some cases school teachers are holding private tuitions. Private tuitions are not a sensitive topic for students, but they are for teachers. Teachers did not volunteer that they were also providing private tuitions.

We inquired about student views on the levelling exam. Students did not mind the levelling test but expressed concerns that it consumes up to two days of classes, when they could be covering material from the course syllabus. The majority of students like Utkarsh and enjoy the more active nature of the lessons. Students report a preference for interactive teaching and group discussion during class. They found Utkarsh lessons and materials to be easier than their regular curriculum, but more fun and engaging. Students expressed interest in linking Utkarsh lessons and methods to their course curriculum/syllabus.

However, some students in focus groups expressed concern that Utkarsh slowed the pace of lessons and thus made it more difficult for them to complete the curriculum. For this reason, students suggested that Utkarsh is valuable for them but would be more desirable during the summer (or before the start of the academic year) such that it does not take away time from covering topics in the official curriculum. During monitoring visits, several headmasters also expressed this same idea, that Utkarsh could possibly be done during vacations or on school holidays. Teachers voiced similar concerns as did students regarding completing both Utkarsh lessons and completing the curriculum. Teachers acknowledged that student learning levels are low and alluded to the lack of foundational skills. Some additional problems faced by teachers include large classes, difficulty disciplining students and excess workloads.

One interesting aspect that came through focus groups is that not all students are told their levels or told during which periods to participate. The mechanics and logistics of Utkarsh are not standardised across all implementing schools. Although students are not always aware of the different phases of Utkarsh, self-reported attendance in the program was high. There was no stigma reported for students attending the Consolidation Camp.

Teachers lauded the programme for its usefulness for low achieving students. However, they shared the students' concerns about delays in the curriculum specific content delivery, and expressed that parents were in agreement about maintaining focus in the curriculum in school. They found Utkarsh to not be beneficial for high achieving students, and suggested delivering Utkarsh in summer or outside of school time. Note that in our preliminary quantitative assessments, we find learning increases throughout the distribution of student ability.

Section 4: Lessons Learned

4.1 Considerations from the Pilot for Utkarsh Implementation

During our data collection, a number of themes emerged in the data and conversations with schools that could be considered for future Utkarsh implementation.

- a. Both students and teachers are concerned that Utkarsh lessons are crowding out regular instructional hours.
- b. The Utkarsh levelling exam could be shortened to make the process easier and less time intensive for teachers. Of course, more validation would need to be done to ensure a valid assessment tool. Students who are absent during the levelling exam are also sometimes unable to have their level assessed.
- c. Students and teachers frequently ignore the placement into camps. Many pupils are attending all phases of the Utkarsh programme, at least partially due to a weak understanding of the programme's particulars and because alternate activities for students not selected for the Foundation Camp and the Supported Learning Phase are not clearly identified.

4.2 Achievement of Pilot Objectives for Proposed Full Study

Overall, we were able to achieve our aims. Below is a brief description of successes and lessons, grouped by the aims set forth in Section 1.

Aim A1: Develop procedures and tools for data collection

1. Group tests are more effective and less expensive: During the pilot, we tested out two types of assessment strategies: short tests administered individually by an enumerator and longer tests administered to groups of students. During the full study we propose to use longer group tests due to fewer challenges in consistent test administration and grading and cost savings.

2. Assessment question and delivery: We piloted and revised different sets of questions to create valid assessment tests that will be used in the study. We also revised testing and grading protocols. For example, language exams will play recordings to measure oral comprehension to avoid challenges in individual enumerator accents and enunciation. We also plan to assess a sample of no more than 20 students per school to reduce the likelihood of cheating. Further, we are exploring strategies to reduce attrition at endline, such as asking headmasters to ensure students attend school that day and allocating money towards tracking absent students.

3. Survey instruments and tools: We piloted and revised questions on the baseline, endline, and monitoring visit survey instruments. We anticipate additional revisions prior to the full study based upon our analysis of survey responses, feedback from enumerators, and input from the focus group discussions.

Aim A2: Validate our student assessments

1. Our evaluation assessments generally closely match with the levelling exam results, suggesting that students are being placed into correct levels and also validating that our assessments effectively capture student learning levels.

Aim A3: Develop and validate tools to measure programme implementation

1. Monitoring Visits: While the school entry and classroom observation protocol was generally deemed sufficient, monitoring visits were disruptive, particularly in control schools. These visits are also relatively more expensive. For the full study we propose to conduct one monitoring visit per school, with visits randomly timed throughout the school year to correspond with key programme periods of Utkarsh.

Aim A4: Develop protocols for tracking students between baseline and endline

1. To successfully match students between baseline and endline, we developed a series of protocols including student name, picture, and other unique identifying information (such as father's name and sibling composition). Our protocols ensured that we were able to re-interview students from baseline who were present during the endline. Attrition will be a key challenge in the full study, and thus we are considering cost-effective ways to track students.

2. Linking students across administrative datasets will continue to be a challenge. We will continue to work with the government and the Transform Trust to collect identifying information that can be used to maximise the use of administrative datasets including potentially the Utkarsh levelling exams and the Class 10 Board exams.

Aim A5: Collect estimates to ensure adequate power and budget

1. Revised power calculations: We collected estimates of key parameters, including attendance/attrition rates, effect size, and the intra-class correlation of test scores. These parameters were useful in ensuring that we have adequate statistical power for the full study with 300 schools.

2. Revised budget estimates: We collected information on the length of time for survey administration, grading and data entry, and revised the budget to ensure that we have a sufficient budget to fully complete all proposed data collection activities.

Aim A6: Identify primary barriers to successful evaluation

1. Sample Selection – JPAL staff and enumerators gave ongoing feedback to the PIs on survey activities, working specifically to identify and address challenges as the evaluation is expanded. The team worked diligently on evaluation details ranging from survey wording and translation to survey implementation to identifying institutional details that would cause issues for the randomised trial. As a result of these efforts, we have revised our sample selection criteria to reduce contamination. We plan to ensure as much as possible that headmasters in the full study do not have schools in different treatment arms. We also listened to our respondents and their concerns. While not captured in our data, we are aware of at least one headmaster in a control school who attended the Utkarsh training. For the full evaluation, we also hope to measure participation in Utkarsh trainings in the event some teachers in control schools receive the training.

Section 5: Conclusion and Next Steps


5.1 Conclusions

Our pilot was successful on many margins. We reached the targeted number of schools, students, teachers and headmasters. We were successfully able to track students over time, although absenteeism is high and will be a challenge for the full study. The data that we collected appears to be high quality and balanced across the treatment and control groups. After data analysis, the broad lessons can be summarised into three points.

First, with several caveats, our results suggest that Utkarsh may increase student test scores on a range of subjects. The magnitude of the point estimates is exciting and comparable to other successful remedial education programmes, although we caution that results in the full study may be smaller. Thus, if we find similar results from the full study, this evaluation will reinforce the need to better tailor pedagogy to learning levels instead of official curriculum based on grade level.

Second, our results suggest that there is support for the Utkarsh programme in the majority of treatment schools. Both students and teachers generally like the Utkarsh lessons as currently implemented; students in particular found the lessons more engaging and interesting. However, concerns about falling behind the official curriculum will likely continue to be a challenge as the programme scales up. Evaluating and considering additional variations to the Utkarsh programme may further broaden its base of support and encourage schools to maximise its potential.

Third, we feel confident that the procedures, tools, and institutional knowledge developed during this pilot provides the framework for conducting a successful large-scale evaluation of this promising programme. Specifically, we feel that this pilot has provided key parameters necessary to ensure an adequately powered and funded evaluation of a three-armed study with 300 schools.



Broadly, given ongoing challenges with poor student outcomes, our results suggest the need for all actors to be more efficient and effective in producing student learning gains. This intervention will test one possible way—the Utkarsh programme—to improve the production of education in the classroom and through improved managerial support. As a result, Utkarsh will provide lessons to both other Indian states and other countries thinking about taking remedial instruction of secondary school students to scale.

5.2 Next Steps

The full study will begin in June 2019, with the baseline activities in July-August 2019 and monitoring visits spread out over the course of August - December 2019 when the programme is being implemented in schools. The final round of data collection – endline – will be conducted in January 2020. These data collection activities will involve interviewing the key stakeholders – headmasters, teachers, and students. We will also conduct focus groups during the 2019-2020 school year. During the 2020-2021 school year, we will follow the students from the full study to measure the effects of the Utkarsh intervention on Class 10 Board Examination scores.

The authors of this report are:

- Sabrin Beg, Assistant Professor of Economics, *University of Delaware*
- Anne Fitzpatrick, Assistant Professor of Economics, *University of Massachusetts Boston*
- Jason Kerwin, Assistant Professor of Applied Economics, *University of Minnesota*
- Adrienne Lucas, Associate Professor of Economics, *University of Delaware*
- Khandker Wahedur Rahman, Ph.D. Candidate in Applied Economics, *University of Minnesota*

Appendix

Appendix Table 1: Baseline summary statistics of headmasters

Variables	(1) Utkarsh		(2) Control		(3) All	t-test (1)-(2)	
	N	Mean/SE	N	Mean/SE	N	Mean/SE	p-value
Headmaster gender (Male=1)	12	.5 (.15)	12	.67 (.14)	24	.58 (.10)	.43
Age (years)	12	51.1 (2.38)	12	51.7 (2.54)	24	51.4 (1.7)	.87
Highest level of education SSS/SHS	12	0 (0)	12	.08 (.08)	24	.04 (.04)	.33
<i>Bachelor degree</i>	12	.42 (.15)	12	.25 (.13)	24	.33 (.1)	.41
<i>Master's</i>	12	.58 (.15)	12	.67 (.14)	24	.63 (.10)	.69
Numbers of years worked as	10	9.6 (3.75)	7	4.71 (1.52)	17	7.59 (2.32)	.25
Numbers of years worked in this	10	2.9 (.64)	7	4 (1.65)	17	3.35 (.76)	.54
How many years did you spend as a subject/class teacher before you became a head	11	16.1 (3.34)	7	24.4 (4.42)	18	19.3 (2.77)	.15
How many days were you absent from	12	.33 (.26)	12	.58 (.29)	24	.46 (.19)	.52
How many teachers in total work here?	12	12.5 (1.62)	12	9.42 (.78)	24	11 (.94)	.1
How many Class 9 teachers does this	12	7.25 (.91)	12	6.17 (.65)	24	6.71 (.56)	.34

Notes: Sample includes 24 headmasters who participated in the pilot baseline survey. Standard errors are clustered at school level. Number of cluster for each variable in each column equals number of observations for that variable in that respective column. The p-values in the last column are from a test that the treatment indicator is zero in a OLS regressions of baseline covariates on an indicator for treatment and using heteroskedasticity-robust standard errors, clustered at the school level.

Appendix Table 2: Baseline summary statistics of teachers

Variables	(1) Utkarsh		(2) Control		(3) All	t-test (1)-(2)	
	N	Mean/SE	N	Mean/SE	N	Mean/SE	p-value
Teacher gender (Male=1)	24	.42 (.083)	23	.48 (.12)	47	.45 (.07)	.68
Age (years)	24	37.3 (2.56)	23	41.5 (2.5)	47	39.4 (1.8)	.25
Highest level of education							
<i>Bachelor degree</i>	24	.67 (.11)	23	.44 (.14)	47	.57 (.09)	.3
<i>Master's</i>	24	.29 (.1)	23	.44 (.12)	47	.36 (.08)	.36
<i>PhD</i>	24	.042 (.042)	23	.04 (.043)	47	.04 (.03)	.98
Numbers of years worked as teacher?	24	11.1 (2.11)	23	16 (2.37)	47	13.5 (1.62)	.13
Numbers of years worked in this school?	23	7.61 (1.8)	23	11.3 (2.16)	46	9.48 (1.41)	.19
How many days were you absent from school last week?	24	.5 (.25)	23	.44 (.16)	47	.47 (.15)	.83
Do your pupils have access to text books for Math? (Yes=1)	23	.96 (.04)	23	.96 (.04)	46	.96 (.03)	1
Do your pupils have access to text books for English? (Yes=1)	24	.83 (.07)	23	.91 (.06)	47	.87 (.05)	.39
How many teacher trainings have you attended this academic year?	24	1.29 (0.7)	23	7 (.29)	47	1 (.15)	.052
Do you currently earn money at other jobs besides being employed as a teacher	24	.08 (.06)	23	.04 (.04)	47	.07 (.04)	.58
Number of clusters	12		12		24		

Notes: Sample includes 48 teachers who participated in the pilot baseline survey. Standard errors are clustered at school level. The p-values in the last column are from a test that the treatment indicator is zero in a OLS regressions of baseline covariates on an indicator for treatment and using heteroskedasticity-robust standard errors, clustered at the school level.

Appendix Table 3: Baseline summary statistics of students

Variables	(1) Utkarsh		(2) Control		(3) All	t-test (1)-(2)	
	N	Mean/SE	N	Mean/SE	N	Mean/SE	p-value
Student has sibling (Yes=1)	322	.92 (.02)	298	.94 (.01)		-0.02	.31
Pupil gender (Male=1)	323	.43 (.1)	299	.43 (.03)		0.00	.98
Age (years)	317	13.5 (.07)	290	13.4 (.05)		0.1	.22
Can your mother read and write? (Yes=1)	323	.77 (.04)	296	.81 (.02)		-0.04	.42
Can your father read and write? (Yes=1)	320	.84 (.03)	298	.85 (.02)		-0.01	.88
How much time does it take you to get to school from your home? (minutes)	323	13.8 (.91)	299	14.4 (.66)		-0.6	.55
What language do you speak at home?							
<i>Odia</i>	321	.93 (.06)	299	.95 (.04)		-0.03	.65
<i>Hindi</i>	321	.1 (.07)	299	.06 (.02)		0.04	.57
<i>Bengali</i>	321	.003 (.003)	299	.02 (.02)		-0.01	.42
Which caste / group do you belong to?							
<i>Scheduled Caste or Scheduled Tribe</i>	291	.3 (.04)	268	.27 (.03)		0.03	.48
<i>Other Backward Caste</i>	291	.4 (.06)	268	.4 (.05)		-0.00	.96
<i>General Caste</i>	291	.3 (.08)	268	.33 (.07)		-0.03	.76
F-stat from test of joint significance of covariates					1.6		
p -value from joint significance of covariates					0.16		
Number of clusters in each arm					12		

Notes: Sample includes 323 students in the Utkarsh schools, and 299 students in the control schools who participated in the pilot baseline survey. Standard errors are clustered at school level. The p-values in the last column are from a test that the treatment indicator is zero in a OLS regressions of baseline covariates on an indicator for treatment and using heteroskedasticity-robust standard errors, clustered at the school level. The p-values from joint significance of covariates are from a test that all the baseline covariates are zero in a OLS regressions of all baseline covariates on an indicator for treatment and using heteroskedasticity-robust standard errors, clustered at the school level.

Appendix Table 4: Monitoring visit summary statistics of teacher survey (questions asked to teachers at both Utkarsh schools and control schools)

Variables	(1) Utkarsh		(2) Control		(3) All	
	N	Mean/SE	N	Mean/SE	N	Mean/SE
Are you currently teaching the lesson that you are supposed to be, according to the official curriculum calendar? (Yes=1)	6	.5 (.22)	6	1 (0)	12	.75 (.13)
Book used						
<i>Official MOE</i>	6	1 (0)	6	1 (0)	12	1 (0)
<i>Utkarsh</i>	6	0.83 (.17)	6	0 (0)	12	0.42 (.15)
<i>Self-made notes</i>	6	0.67 (.21)	6	.5 (.22)	12	0.58 (.15)
Pace of progress in class 9 syllabus						
<i>Ahead</i>	6	0 (0)	6	0.17 (.17)	12	0.08 (.08)
<i>Behind</i>	6	.5 (.22)	6	0 (0)	12	.25 (.13)
<i>On schedule</i>	6	.5 (.22)	6	.83 (.17)	12	.67 (.14)

Notes: Sample includes 12 teachers who were surveyed in the pilot monitor survey. Standard errors are clustered at school level

Appendix Table 5: Monitoring visit summary statistics of teacher survey (questions asked to teachers at Utkarsh schools only)

Variables	N	(I) Utkarsh Mean/SE
Have you attended Utkarsh training? (Yes=1)	6	.83 (.17)
Did you receive a teaching handbook for the Utkarsh training? (Yes=1)	6	1 (0)
Which phase of the Utkarsh program is currently in place at your school? SLP	6	1 (0)
Who are supposed to participate in FC phase? <i>All students</i>	6	.17 (.17)
<i>Inception level students</i>	6	.5 (.22)
<i>Class 3 level students</i>	6	.83 (.17)
<i>Class 5 level students</i>	6	.83 (.17)
<i>Class 8 level students</i>	6	.17 (.17)
Who are supposed to participate in SLP phase? <i>All students</i>	6	.67 (.21)
<i>Inception level students</i>	6	0 (0)
<i>Class 3 level students</i>	6	.17 (.17)
<i>Class 5 level students</i>	6	.33 (.21)
<i>Class 8 level students</i>	6	.17 (.17)
What do you think about the following statement-“I enjoy teaching Utkarsh lessons in the classroom, relative to the standard curriculum”		
<i>Neutral</i>	6	.17 (.17)
<i>Agree</i>	6	.33 (.21)
<i>Strongly Agree</i>	6	.5 (.22)

Notes: Sample includes 12 teachers who were surveyed in the pilot monitor survey. Standard errors are clustered at school level

Appendix Table 6: Monitoring visit summary statistics of teacher survey (questions asked to teachers at Utkarsh schools only)

Variables	N	(1) Utkarsh Mean/SE
What do grade-level students do during Utkarsh camp?		
<i>Attend the lesson as well</i>	6	.83 (.17)
<i>Sit idle</i>	6	.33 (.21)
What do grade-level students do during Utkarsh camp?		
<i>Attend the lesson as well</i>	6	.83 (.17)
<i>Sit idle</i>	6	.33 (.21)
In your school, which Class 9 students attended the Utkarsh Foundation Camp lessons?		
<i>All students</i>	6	1 (0)
In your school, which Class 9 students attended the Utkarsh SLP lessons?		
<i>All students</i>	6	.83 (.17)
<i>Don't know</i>	6	.17 (.17)
Today, did you follow the scheduled Utkarsh lesson for the day's class? (Yes=1)	6	.5 (.22)
Did your school deliver the Foundation Camp? (Yes=1)	6	1 (0)
How hard has it been to do Utkarsh lessons and activities?		
<i>Somewhat hard</i>	6	.5 (.22)
<i>Somewhat easy</i>	6	.33 (.21)
<i>I don't follow the lesson</i>	6	.17 (.17)

Notes: Sample includes 12 teachers who were surveyed in the pilot monitor survey. Standard errors are clustered at school level

Appendix Table 7: Effect of Utkarsh treatment on endline test scores (in control-group SDs)

	Math (1)	English (2)	Odia (3)	Science (4)
Treatment	0.270 (0.137)	0.280 (0.128)	0.355 (0.139)	0.285 (0.138)
Scoring Method	IRT	IRT	IRT	IRT
Test Type	Pooled	Pooled	Pooled	Pooled
IRT Weights	Pooled	Pooled	Pooled	Pooled
Observations	441	441	438	439
Adj. R-Squared	0.553	0.586	0.434	0.263

Notes: Sample includes 230 students in the Utkarsh schools, and 211 students in the control schools who participated in the pilot endline survey. Robust standard errors in parentheses, clustered by school. Columns 1 and 2 control for an indicator for the type of test (Adaptive or Group). IRT weights refers to the sample of test scores used to estimate the IRT parameters, i.e., group test weights for students who took the group test and adaptive test weights for students who took the adaptive test. All regressions control for the student's within-school percentile of the baseline test score, age-by-gender categorical interactions, mother's and father's literacy, and whether the student has any siblings.

References

- Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukherji, S., Shotland, M. & Walton, M. (2017). Mainstreaming an Effective Intervention: Evidence from Randomized Evaluations of “Teaching at the Right Level” in India. *Journal of Economic Perspectives*, 31(4):73-102.
- Duflo, A., Kiessel, J., & Lucas, A. (2018). Scaling successful programs in a new context: The Teacher community Assistant Initiative. University of Delaware Working Paper.
- Das, J., & Zajonc, T. (2010). India shining and Bharat drowning: Comparing two Indian states to the worldwide distribution in mathematics achievement. *Journal of Development Economics*, 92(2), 175-187.
- National University of Educational Planning and Administration. (2016). *School Education in India Flash Statistics 2015-16*. New Delhi: NUEPA.
- Reserve Bank of India. (2018). *Handbook of Statistics on Indian States 2017-18*. New Delhi: RBI.
- Reserve Bank of India. (2017). *Handbook of Statistics on the Indian Economy 2016-17*. New Delhi: RBI.

Kusuma Trust UK

5th Floor | 55 New Oxford Street | London | WC1A 1BS
T: +44 (0)20 7420 0650 | www.kusumatrust.org