Teacher Flexibility and School Productivity: Remedial Secondary Education in India

Sabrin Beg, Anne Fitzpatrick, Jason Kerwin, Adrienne M. Lucas,

and Khandker Wahedur Rahman*

September 13, 2024

Abstract

In public sector organizations, there is an inherent tension between allowing flexibility at the point of service to adapt to local stakeholder needs versus rigid harmonization to ensure uniform service delivery. Through a 300-school RCT in Odisha, India, we compare a rigid vs. flexible version of a remedial learning intervention, along with a control group. Both interventions increased students test scores by 0.11SD, about 60 percent of a year of learning, with gains throughout the learning distribution. Despite concerns that allowing teachers flexibility to adjust their lessons would lead to shirking and ineffective teaching practices, we find no evidence that the quality of implementation was affected by the flexibility: teaching practices improved symmetrically by 0.3 standard deviations in both arms. Remedial education is an effective use of class time, did not crowd out grade level mastery, and allowing teachers flexibility to adjust lessons to student needs may be beneficial.

*Beg: University of Delaware and J-PAL. Fitzpatrick: The Ohio State University. Kerwin: University of Washington, IZA, and J-PAL. Lucas: University of Delaware, J-PAL, and NBER. Rahman: University of Oxford and BIGD, BRAC University. All study activities were approved by the following Institutional Review Boards: IFMR, University of Minnesota, University of Delaware, and University of Massachusetts Boston. This evaluation would not be possible without our partners at Transform Schools, People for Action and the School and Mass Education Department of Odisha, including Pankaj Sharma, Bindiya Nagpal, Christine Oliver, and Kartik Sahu. We thank Rebecca Thornton, Sandra Sequiera and seminar participants at Bentley University, Boston University, IFPRI, Lafayette College, Universite Laval, the University of Delaware, University of Nevada Reno, Wellesley College, the University of Maryland, Miami University, as well as our discussants at the following conferences: AEFP, AFE, CSAE, NEUDC, and PAA for their comments. We also thank the JPAL-South Asia field operations staff, and all of our respondents for their time, cooperation, and insight. Soumya Das provided excellent research assistance. Funding for this project was provided by the UK Foreign, Commonwealth & Development Office, awarded through J-PAL's Post-Primary Education Initiative, and the Kusuma Trust UK. AEA RCT #AEARCTR-0004138

JEL Codes: H40, I21, I2, J24, J28, M50, M53, M54, O15, O43.

Keywords: randomized trial, remedial education, secondary school, India

1 Introduction

Public sector organizations in developing countries are often rigidly coordinated bureaucracies where front-line civil servants (e.g., teachers, health workers) are directed to operate according to prescribed, uniform rules(Afridi et al., 2017; Bandiera et al., 2020; Banerjee et al., 2021; Muralidharan and Singh, 2020). This organizational design often ignores local conditions, limits the role of the front-line civil servants in decision-making and may contribute to observed substandard service delivery and specifically poor learning outcomes. Yet such rules might be necessary in the absence of other incentives for good job performance, or informational and capacity constraints.

Indian government schools are a classic example of highly regimented public-sector organizational structure with poor service delivery outcomes. Teachers (i.e., front-line civil servants) are directed to follow and complete the grade-level curriculum on a strict timetable decided by the state education office even though they are aware that more than half of students are multiple grade levels behind the curriculum (Das and Zajonc, 2010; Muralidharan et al., 2019; Muralidharan and Singh, 2020). The current system of standardized teaching approaches and uniform implementation across a state is designed to create high expectations of teachers and students, forestall issues of the moral hazard of teacher shirking, and decrease planning, monitoring, and coordination costs. Centralized service delivery may also be practical given concerns about teacher competency, especially at the secondary school level (Chaudhury et al., 2006; Duflo et al., 2012, 2015)(Beteille et al., 2020; Bold et al., 2017b,a, 2018). However, there is some evidence that teachers ignore these dictates, by either shirking or teaching according to their own preferences anyway (Muralidharan and Sundararaman, 2011). In this context, formally decentralizing decision-making to the school, or to the classroom may have unrealized benefits. Letting teachers decide what and how to teach may increase the potential for creativity and intrinsic motivation (Pearson and Moomaw, 2005; Skaalvik and Skaalvik, 2014). Clearly the current approach to managing teachers in India is not working well—but why? Is regimentation bad, taking discretion

away from local decision makers? Or is the problem that what is regimented is inappropriate? In this paper, we address fundamental questions relevant to teaching worldwide: what should be taught in a classroom, and who should make that decision?

To do this we use data from a 3-armed randomized controlled trial across 300 secondary schools in Odisha, India that tested the effect of of a remedial education program that replaced standard curriculum lessons with remedial content, and also changed the degree to which teachers were allowed flexibility over classroom content. Specifically, we partnered with existing school personnel, the Odisha Department of School and Mass Education, and the Indian non-governmental organization Transform Schools, People for Action to test whether the *Utkarsh* ("excellence") program is more effective than the status quo ("the curriculum"). In the first Treatment group, teachers in 100 schools were provided remedial lessons through training, materials, and an adjusted time table that explicitly mandated that teachers engage in the remedial activity instead of the usual curriculum. *Utkarsh* lessons took place during school in a bootcamp style: multiple hours for a limited number of days were set aside for the remedial lessons. These were not additional schooling hours and directly displaced curriculum-level lessons. In the second Treatment group, teachers in 100 schools were randomly assigned to receive the *Utkarsh* program with additional freedom to implement the program with more flexibility based upon their assessment of students' needs. This *Flexible Utkarsh* arm gave teachers explicit permission to re-focus class time on either previous *Utkarsh* content or grade-level material, depending on what they felt was a priority. With this study design we are able to test and separate the relative learning gains from a) directing teachers to teach a potentially more appropriate classroom content; b) allowing teachers more control over classroom activities.

We report four main findings based on one school year of implementation. First, students were substantially behind grade level—an average Grade 9 student was 4 to 5 grades behind in English, in math, and Odia (the local language) with substantial within-school heterogeneity. In contrast to work by Djaker et al. (2022), teachers were largely aware that students were

substantially behind grade level, with some uncertainty in their estimates.

Second, both treatment arms improved student learning. In both treatments, learning improved by about 0.1 standard deviations (SDs), equivalent to a 58 percent increase in learning compared to the control group, and increased the likelihood of achieving grade 3 and grade 5 competency by between 3 to 8 percentage points, depending upon subject. However, the intervention did not bring students up to grade 9 level mastery of the material. The intervention neither crowded out grade level knowledge nor harmed grade-level students during the period of the intervention, both of which are key concerns about introducing remedial lessons (Zhao, 2018; Figlio and Page, 2002). Overall, this program was also cost-effective, costing 0.95 SDs per $100, when implemented at a 200-school scale.

Third, implementation quality was high in both treatment arms despite no additional incentives—teachers implemented the program in line with the stated rules. We observe no substantial increase in shirking in the *Flexible Utkarsh* arm, a key concern of allowing more decision-making at the point of service. Teachers usually taught *Utkarsh* during the designated times, but not always the exact lesson prescribed for that day—teachers in both treatment arms modified the program, but stuck with remedial lessons instead of reverting to the standard curriculum. These results suggest that teachers already adapt classroom activities for their students' specific needs, and also imply a revealed preference for additional support (such as teaching and learning materials) to modify classroom activities to better fit student learning abilities. Finally, at the end of the treatment, teachers had more accurate beliefs about their students' learning levels. Therefore, even though they believed that they and their students benefited from the program, teachers' assessment of their students' knowledge was lower than teachers in control schools.

This paper makes several contributions to the academic and policy literature. First, these results are highly relevant to education policy: this program has been scaled up both across the original state (Odisha, population 40 million) and also a new state (Karnataka, population 60 million). In addition to the policy relevance of the results, this paper makes three

major contributions to the economics literature. First, we show that there are substantial learning gains from acknowledging local conditions in organizational decision making: remedial instruction provided by existing government teachers at the secondary level increases student learning by 58 percent. Previous research has shown that remedial instruction delivered via tracking, pulling students out of class, or adding additional instruction time outside the school day increases student learning for primary school students (Banerjee et al., 2007, 2010, 2017; Duflo et al., 2011, 2020; Lakshminarayana et al., 2013).[1] In contrast, limited evidence exists for remedial instructional at the secondary school level, with interventions either outside of the school day with tutors from outside the system (Muralidharan et al. (2019)) or within school, but focusing on grade level material (Beg, Lucas, Halim and Saif, 2019). One notable exception is Lavy and Schlosser (2005), who evaluate a multi-year government-led remedial education program that occurred outside of school hours, targeting low performing high schools in Israel using a difference-in-difference model; they find that the program improves matriculation rates. However, whether within-school remedial education is a beneficial use of class time is an outstanding question. The challenges teachers face at the secondary school level are substantially different than that of primary school. For example, because students are often promoted despite not mastering grade-level competencies, secondary school classrooms are likely to be substantially more heterogeneous. As a result, remedial lessons are more difficult to target and could be less effective. Overall, at the secondary school level, it may not be optimal to invest in students who are substantially behind; it may simply be too late. On the other hand, some students *are* at the level of the curriculum; introducing remedial lessons to all may actually harm students by crowding out grade-level material. We provide new evidence that remedial education provided by existing personnel, during the school day, is an efficient use of existing school resources at the secondary school level.

Second, we provide new evidence that allowing decision-making at the point-of-service

[1]See Appendix Figure A1 for a summary of the recent literature on remedial education in the form of TaRL (Teaching at the Right Level) conducted in low-income settings.

may improves the functioning of highly rigid bureaucracies (or at least do not harm). Much of the research in improving public sector service delivery focuses on providing incentives to service providers (Rasul and Rogger, 2018; De Ree, Muralidharan, Pradhan and Rogers, 2018; Duflo, Dupas and Kremer, 2011; Duflo, Hanna and Ryan, 2012; Duflo, Dupas and Kremer, 2015; Glewwe, Ilias and Kremer, 2010; Muralidharan and Sundararaman, 2011; Barrera-Osorio and Raju, 2017; Brown and Andrabi, 2020; Rasul, Rogger and Williams, 2018) or empowering community members to register complaints (Bjorkman and Svensson, 2009; Duflo et al., 2012). A growing literature emphasizes that making highly rigid bureaucracies less rigid may improve service delivery by increasing autonomy among front-line civil servants (Bandiera, Best, Khan and Prat, 2020; Bloom, Lemos, Sadun and Reenen, 2015), though it is difficult to find optimal practices in a hierarchical bureaucracy (Banerjee et al., 2021). Whether these reforms would improve educational quality, however, is unclear. In Kenya, private schools that enforced highly regimented (i.e., scripted) lessons substantially improved learning, suggesting that there many be currently too much autonomy in schools (Gray-Lobe et al., 2022). Piper, Sitabkhan, Mejia and Betts (2018) similarly find that guided lessons improve student learning, but the majority of teacher modifications decrease lesson quality. We find Treatment-group teachers replaced curriculum lessons with remedial lessons as instructed and when given the opportunity to deviate back to the grade level curriculum, stuck with the remedial lessons, which was at a more appropriate learning level for their students. Our results contrast to Gray-Lobe et al. (2022) and Piper, Sitabkhan, Mejia and Betts (2018), and show that allowing teachers additional flexibility is not worse when their lesson content is appropriate. We interpret these results through a conceptual framework closely related to the model of Dessein (2002) in the field of organizational economics that outlines the conditions under which delegation of authority to local decision-makers is optimal. Teachers already deviate from centrally-set dictates; therefore, there may be benefits from explicitly giving authority to teachers to do what they see fit in the classroom.

Third, we use our rich data to analyze the mechanisms of our results, and provide new

evidence on teacher decision-making in the classroom. Thus, our study has lessons for how organizations can leverage civil servant preferences to improve outcomes. There are three main reasons why teachers might teach content that is at an inappropriately high level: 1) they feel that they have to complete the curriculum due to bureaucratic or parental pressure; 2) they do not know the degree to which students' learning levels are out of sync with the curriculum, or 3) remedial lesson teaching preparation is costly in terms of effort. Our results are most consistent that teachers are willing to teach remedial education, and also recognize its value; instead, the most binding constraint is the costly effort involved with deviating from the curriculum and preparing lessons that are at the level of student. Previous interventions that only provided materials (Glewwe et al., 2004) or materials and training in primary schools did not increase learning (Banerjee et al., 2017; Duflo et al., 2020). By contrast, we find that teachers in our setting are interested and willing in engaging in new directives to increase student learning, with additional support.

2 Background

Our study takes place in Odisha, a relatively poor state in eastern India where about 32.6 percent of the population is below the national poverty line (NITI Aayog, 2021).Many secondary school students are first-generation learners and about 40 percent of enrolled students are either scheduled caste or scheduled tribe. In India, pre-tertiary education schooling is primary school (Class 1 to Class 5), middle school (Class 6 through 8), lower secondary (Classes 9 and 10), and higher secondary (Classes 11 and 12). Teachers in lower secondary, the focus of this study, are subject teachers, teaching the same subject to Class 9 and 10 students in separate sections, and sometimes teaching multiple subjects across the grades depending on the school size. As is typical in the Indian education system, schools in Odisha emphasize teacher-focused instruction, have many below grade level students, and heterogeneous learning levels in the same classroom. The typical class period involves lecture-based pedagogy that strictly adheres to the official curriculum with limited pupil participation or

deviation for students at learning levels below the official levels. At baseline, 95 percent of headmasters consider adhering to the curriculum to be an important component of their job. At the end of the school year, headmasters conduct a review of student participation, concluding assessments and Class 9 exam results to assess student learning progress in relation to the curriculum.

Students must pass standardized, district-run Board exams at the end of Classes 10 and 12 to continue their education. The marks on the Board exam determine where and what field of future study students can undertake and are intended to ensure adequate student grade-level competency. However, Das and Zajonc (2010) find that approximately 50 percent of the enrolled secondary school students in Odisha fail to meet a basic benchmark of mathematical knowledge, leading to concerns of teaching quality.

The school year in Odisha starts in April, has a break from early May to mid June, and ends in March. Our study follows students from the beginning of lower secondary school (Class 9) through their first year in higher secondary (Class 11). The intervention occurred during Class 9 only.

3 The Utkarsh Program

The *Utkarsh* program provided Class 9 teachers training, teaching and learning materials, and a designated schedule to follow to deliver this content. The program was a collaboration between the Odisha School Education Programme Authority (the authority within the Department of School and Mass Education, or SMED, responsible for education) and Transform Schools, People For Action, a large Indian NGO. It focused on Odia, English, math, and science and was designed to improve learning outcomes for students who were below grade level. All instruction took place within the existing school day with existing teachers.

Our study covers two versions of *Utkarsh* – *Standard Utkarsh* and *Flexible Utkarsh*. We describe each below.

Standard Utkarsh

The program involved teacher training, teaching and learning materials, and a change in the allocation of time during the school day. Subject teachers of English, Math, Odia, and Science and the school headmaster were invited to a one-week training session in August of the 2019-2020 school year. At these sessions, participants learned how to use Utkarsh teaching and learning materials to implement a more effective teaching practice. The materials were remedial and covered the content that should have been covered in Classes 3 through 8, divided into specific Class-specific phases. The training also emphasized collaborative and student-centered active learning, including students sitting and working in groups. Daily outlines of topics to cover and accompanying student worksheets were provided; lessons were not scripted. Appendix A contains additional details.

Upon returning to school, teachers were to test all students in Class 9 to determine their learning levels based on provided rubrics. Students were categorized as inception (below Class 3), Class 3, Class 5, and Class 8 or above. The level of each student determined in which phase of the lessons the student would participate: Foundation Camp, Supported Learning Phase, or Consolidation Camp. The program instructed teachers to provide alternative activities for students who were did not need remediation, although in practice students who did not need remediation participated in the same lessons, in the same classroom, for all parts of the intervention .

Foundation Camp (FC): FC was for the students who initially tested at the Class 5 level or below and was designed to support the learning of foundational concepts and skills. This phase was 4 hours per day for 18 days, for a total of 72 hours of instruction.

Supported Learning Phase (SLP): SLP targeted all students who tested at the start of the year below a Class 8 level, about 90 percent of students in our sample. SLP further developed the FC concepts at a higher level and with more advanced skills, moving students from Class 5 to Class 7 level material. Teachers were to adhere to these lessons for 3 hours per day for 45 days, for a total of 135 hours of instruction.

Consolidation Camp (CC): The final phase, CC, included all students and focused on grade-level material in preparation for the Class 9 annual examinations. CC was 3 hours per day over 6 days, for a total of 18 hours of instruction.

At the end of the CC phase, approximately four months after the start of the program, teachers again assessed all students on their learning.

Flexible Utkarsh

Teachers and head masters received the all of the training and materials as the *Standard Utkarsh* version described above. They were also instructed to implement both Foundation Camp and Consolidation Camp as above. However, during SLP, teachers could follow the *Standard Utkarsh* plan as above or they could exercise their own discretion in planning the material and content during the 3 hours per day of SLP time that occurred over 45 days. They could spend more time on a particular SLP topic, skip SLP topics if their students did not need them, or use the time for standard curriculum instead of remedial SLP topics. They were encouraged to complete their *Flexible Utkarsh* SLP plan each week with the topics that they planned to cover for the week, whether copying from the standard plan or deviating. The training requested teachers to cover at least 50 percent of the material from the standard SLP curriculum, but could do so in any order and devoting less time to these lessons based on their local expertise about their students.

4 Conceptual Framework

We use as a starting point the seminal model of Dessein (2002) in the economics of organizations. This model outlines the conditions under which it is optimal for a principal (for example, company owner, or, in this case, a bureaucrat) to delegate decision-making authority regarding to a subordinate (in this case, a teacher). We apply this model to the decision of who should make the decision of what to teach in a particular classroom. While a large number of models in political science outline the trade-offs of decision-making authority versus delegation in bureaucracies– such as (Gailmard, 2002) and (Jo and Rothenberg, 2014)

—there are several aspects of the Dessein (2002) model are particularly relevant to modeling educational choices, and particularly whether to prescribe a centrally-set curriculum or instead allow for point-of-service modifications. First, we assume that bureaucrats (the "principals") have a preference to maximize student learning, and have the authority to decide instructional content, such as teaching a given curriculum or instead remedial education. However, they lack key information regarding local conditions: specifically, they do not have sufficient information regarding the distribution of abilities of the students at a particular school. Thus, they must make a decision about curriculum without knowing what specific outcome would be maximize learning. Teachers (the "agents") have this information, but may have biased preferences as compared to the bureaucrats. For example, they may wish to shirk, or have a preference for a certain pedagogy. Furthermore, the bureaucrats cannot design a contract with teachers to elicit the private information regarding the actual abilities of their students and act upon that information. Instead, bureaucrats may either: 1) take the teachers' reports on the local conditions and make the decision regarding what content the teacher should teach; or, 2) allocate decision-making power on what content to teach to the teacher. In other words, the bureaucrat must decide whether the bureaucracy should make the decision regarding instructional content, after consulting with civil servants; or, they can delegate the decision.

The bureaucrat is considering various policies over classroom teaching; these different curriculum guidelines are captured by $y \in \mathscr{R}$. Each of the various curriculum offerings that generate to them a private benefit

$$U_B(y, m)$$

where m is a random variable with density $f(m)$, bounded over some range. The bureaucrat does not observe m. By contrast, the teacher does observe m. However, teacher's utility is $U_T(y, m, b)$ where b is an additional parameter that captures the extent to which the bureaucrat's preferences differ from that of the teacher's. Following Dessein (2002) we use the same form of the utility function for the bureaucrat:

$U_B(y, m) = U_B(m, m) - \lambda(|y - m|)$

where λ is a function with a positive second derivative and $\lambda(0) = 0$. We note that the bureaucrats utility is maximized when $y = m$. By contrast, the teacher's utility is maximized when $y = m + b$:

$U_T(y, m) = U_T(m + b, m) - \lambda(|y - (m + b)|)$

Because only the teacher knows the observed value of m, the bureaucrat has two choices: delegate authority (i.e., allow the teacher to choose y), or else rely upon a possibly biased teacher report of the value of m before choosing y.

Proposition 1: Teachers will choose to implement *Utkarsh* program if their preferences are sufficiently close to that of the bureaucrat.

Dessein (2002) shows that, together with general assumptions on the structure of the private information, delegation (i.e., allowing the teacher to choose y) is optimal if and only if the bias b is smaller than some cutoff value. Thus, delegation is optimal for a range of values, and does not require that the preferences of the teacher and the bureaucrat are in alignment; the teacher just cannot be sufficiently biased towards their own preferences to distort the bureaucrat's best option. Note how this dynamic holds even though the bureaucrat cannot directly control what action the teacher ultimately chooses in the classroom.

Applying this model to the context of Indian education leads to two fundamental questions. First, what are the preferences of the bureaucrat? Second, what are the preferences of teachers? Mapping this conceptual framework to our experiment, we assume that bureaucrats ultimately want to maximize student learning, as opposed to (for example) having a preference for the curriculum above all else. Specifically, we assume that bureaucrats want to have teachers implement remedial education instead of the curriculum for part of the day. This assumption is somewhat innocuous, as this maps directly to our experimental design. However, teacher preferences– and the degree to which they agree with the preferences of the bureaucrat– is ultimately an empirical question. Shirking is a concern, both on the extensive and intensive margin. For example, (Chaudhury et al., 2006)find that as many as 19 percent

of teachers are absent at any given time; other studies have similarly found high rates of both absenteeism and low levels of effort on the job (Duflo et al., 2012). Furthermore, teachers' preferences on what to teach in the classroom may be different than that of policymakers and other stakeholders. For example, if policymakers dictate teachers begin implementing remedial instruction, they may or may not do so.

We now consider an extension to the model to better understand how delegation interacts with both teacher and policymaker preferences. We assume that the bureaucrat can provide training, teaching, and learning materials to facilitate remedial instruction (i.e., can adopt *Utkarsh*). The bureaucrat thus lowers the effort costs of teachers to doing activities to help ensure that their preferences are in line with that of the bureaucrat. We include an additional term b_P , representing the support that teachers receive from the bureaucrat, making their utility as follows:

$U_P(y, m) = U_P(m + b_P, m) - \lambda(|y - (m + b_P)|)$

$b_P = (1 - k)b$, where k $\in$ (0,1)

With this additional term, the teacher's preferences are closer to that of the bureaucrat.

Proposition 1: Teachers may be reluctant adopt point-of-service modifications as they have different preferences as the bureaucrat. However, they will choose adopt *Utkarsh* in their classrooms if they are provided teaching and learning materials that help offset the costs of instituting reforms in their classroom.

Proposition 2: With an explicit delegation of authority to teachers, delegating decision-making to the teacher can be optimal as long as the bias is at an intermediate level.

Similar to before, if the bias is low, then the bureaucrat prefers to allocate decision rights to the teacher; if the level of bias is too high, then the bureaucrat prefers to not delegate decision-making at all. We now test this theory with a randomized evaluation that decentralized decision-making power regarding how classroom time was allocated, testing whether the status quo could be improved with increased delegation.

5 Empirical Strategy

The primary conceptual difficulty in assessing the effect of remedial learning on student outcomes is the typical correlation between remedial instruction and students' learning level or other school characteristics. To overcome this difficulty, we randomly assigned each of the schools in our 300 school sample into one of three groups: 1) *Standard Utkarsh*; 2) *Flexible Utkarsh*; and 3) control, i.e., business as usual.

We estimate the impact of the two variants of *Utkarsh* using the following equation:

$$y_{ist} = \alpha + \beta_1 StandardUtkarsh_s + \beta_2 FlexibleUtkarsh_s + \delta' X_{ist} + \varepsilon_{ist} \quad (1)$$

where y_{ist} is the outcome of interest for respondent i in school s at time t. $StandardUtkarsh_s$ and $FlexibleUtkarsh_s$ are dummy variables indicating the randomly assigned treatment status of school s. These indicators are mutually exclusive with the control group as the omitted category. X_{is} is a vector of control characteristics, including the baseline value of the outcome variable (as appropriate), the wave of survey (if the outcome is measured at multiple waves), and strata, day of the week, and week of the year fixed effects. [2] Standard errors, ε_{ist}, are clustered at the school level.

Our coefficients of interest are β_1, the effect of *Standard Utkarsh* relative to the control group, and β_2, the effect of *Flexible Utkarsh* relative to the control group. The difference between $\beta_1$and β_2 is the difference in the effects of the two interventions.

Our primary outcomes of interest are student test scores at the conclusion of the intervention and on program implementation. To understand the mechanisms behind test score changes, we also estimate the effect of the interventions on teacher classroom behavior and practices and teachers' perceptions of students. We also analyze the effects on longer-term measures of the program, namely the Board marks.

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

6 Sample Selection, Randomization, and Data

6.1 Sample Selection and Randomization

To arrive at our 300 school sample we started with an administrative list of all 711 secondary schools in Jajpur and Dhenkanal districts in Odisha State, India. Schools that did not report any students enrolled in Class 9 were eliminated, leaving 348 villages with at least one secondary school. To minimize contamination, we randomly selected one secondary school from each of these villages. We randomly ordered these schools and directly confirmed with each one that it used the the official state language (Odia), was governed by the SMED and not the Scheduled Caste-Scheduled Tribe Development Department, had students enrolled in Class 9, and was not a school for students with special needs (e.g., deaf or blind students) until we reached 300 schools that passed the screening criteria. We placed each of the 300 schools that passed the screening test into one of 46 strata based on district, average pass rate on the prior year's Class 10 board exam, total Class 9 enrollment, teacher to student ratio, and distance to the district headquarters. Within each strata, we randomized an equal number of schools into the three treatment conditions, resulting in 100 schools in each of the three treatment arms. Figure 1 shows this study design.

6.2 Data Collection

We conducted four waves of data. Three were collected during the the year of implementation (academic year 2019-2020): a baseline survey, an unannounced monitoring visit when schools should have been engaged in *Utkarsh*, and a full follow-up at the conclusion of the intervention. The year after the intervention, we conducted an additional followup survey via phone. We augment these data with administrative data from 2021.

Baseline

The baseline surveys took place in July and August 2019, near the start of the school year but after the summer break and prior to the implementation of Utkarsh. We conducted

demographic and background information from the school headmaster, teachers of the four *Utkarsh* subjects, and sample students; data about the school's infrastructure; and invigilated exams in Odia, math, and English. See Appendix B for additional test construction details. These exams were separate from the Utkarsh leveling exams, which teachers had not yet conducted at the time of our baseline exams.

Monitoring Visits

Between September and November 2019 we conducted one monitoring visit at each school. During these three months schools should have been implementing FC and SLP. We randomly assigned each school into receiving their visit during one of three monitoring visit phases: FC, early SLP, and late SLP. We block randomized assignment to monitoring visit phases by district and study arm. Monitoring visits in FC-phase schools took place when the FC phase of Utkarsh was going on. Similarly, in early-SLP-phase schools, monitoring visits took place immediately after SLP begun, while in late-SLP-phase schools we conducted monitoring visits late in the SLP phase of Utkarsh. During each visit, enumerators arrived unannounced and recorded the attendance of the head master, teachers, and baseline students. Head masters and teachers responded to questions about program take-up and implementation. We also conducted classroom observations.[3]

First Follow-up

We conducted the first follow-up from December 2019 to February 2020 at the end of the intervention, slightly overlapping with the conclusion of the intervention in December 2019.[4] Students responded to a short student survey that included a question about their board exam registration and registration number and completed subject exams in Odia,

[3]Classroom observations occurred during the first period of the day. Enumerators sat in a classroom for one period and collected data on teacher behavior and presence, student behavior, and the use of teaching and learning materials. Monitoring visits occurred in 298 schools; 2 schools did not consent.

[4]Treatment schools should have still been still implementing the CC at the start of the fieldwork in December 2019. We randomly selected 9 strata to visit during December 2019, visiting all treatment and control schools in those strata. We visited 60 schools in December and the remaining 240 in January and February. The rationale for splitting the endline was because it was not possible to complete data collection prior to schools beginning their preparations for end-of-term exams. During this time, it was unlikely schools would grant us permission to enter and conduct surveys. The treatment effect for test scores are the same for early versus late endline.

English, mathematics, and science. These tests were similar to those at baseline, but included additional more challenging questions and a science exam. We sought to interview and invigilate exams for all baseline students. Our analysis sample is all 5,448 students that completed both the surveys and assessments. During this follow-up visit, we also conducted surveys of teachers, and the school headmaster. The teacher survey asked teachers about their experience, autonomy, Utkarsh implementation, work load, and perception about Utkarsh. We also administered a competency test on English and math to teachers to test their knowledge on respective subject matters.[5] The headmaster survey covered a range of topics from information on their school and characteristics, to their personal background and school management practices. We confirmed the board exam registration information for each student in the sample with their head master.

To maximize response rate for this follow-up, we followed DiNardo et al. (2021) and randomized the intensity of our mop-up visits at all schools to survey respondents (student, teacher, and headmaster) who were absent during the follow-up visit.[6] Specifically, we conducted second mop-up visits in a random subset of schools where students remained absent during the first mop-up visit. We implement Lee bounds (Lee, 2009) to address potential biases of treatment effects due to non-random coverage of respondents in the follow-up.

Second Follow-Up Survey

We conducted a second follow-up survey via phone in December 2021, two years after the end of the program, to measure the impact of the program on longer-term school enrollment and the transition to additional schooling or work.[7]

Administrative Board Marks

[5]Creating a consistent and comparable sample of teachers across the study waves was challenging due to teacher vacancies and guest teachers (who may not work at the school each day of the week). Hence, we attempted to survey the same teachers over time, adding teachers as necessary and collecting demographics as they were added.

[6]Despite mop ups, in some cases we were unable to record student board exam registration for all students because the headmaster did not have time to access or could not access digital copies of the Class 10 board registration.

[7]This survey successfully reached 1,255 of the students from baseline (23 percent). This rate is similar across the treatment arms.

As additional evidence on the longer-term effect of the program, we acquired the Class 10 Board Marks for the entire cohort of students in our sample schools. Because of the Covid-19 pandemic, the Board Marks for these students were calculated in an atypical way. The end of the Class 10 year was May 2021, and due to the COVID-19 pandemic, the May 2021 Board Exam was canceled. Instead, students received a Board Mark based on a weighted average of their in-class Class 9 scores (40 percent) and their in-class Class 10 scores (60 percent). Teachers assessed students in Class 9 and 10 without knowing that they would be used as Board Marks. As a result of some students objecting to this grading scheme, an optional formal Board Exam was eventually administered, which approximately 5 percent of Class 10 students completed. The Board Marks recorded in the administrative data are the maximum of the school-based weighted average and the Board Exam. Unfortunately, the administrative data do not denote whether the student sat the formal Board Exam or whether the final Board Marks are from the Exam or the school-based weighted average.

6.3 Summary Statistics and Baseline Balance

Appendix Table A1 and A2 show that randomization successfully created three groups with balanced characteristics at baseline at the student, school, and teacher level. Approximately half of the students in the sample are girls, and the average age is approximately 13 years. Nearly two-thirds (61 percent) of students belong to either a scheduled caste, a scheduled tribe, or other backwards caste, the disadvantaged minority groups in India. About 15 percent of students in our sample have illiterate parents. Slightly less than half of the teachers are female (48 percent), and the average teacher is 42 years old. Approximately one-third (35 percent) of teachers have a teaching certificate, and the average teacher has 16.5 total years of experience. Teachers report very little absenteeism from work. Teachers report spending about 21 hours each week preparing lessons and grading. Teachers believe that approximately 60 percent of their students will pass board exams on their first try.

About 22 percent of headmasters were women. Our headmaster survey also indicates

the dearth of autonomy in schools to adjust curriculum to meet the learning-level need of students. About 77 percent of headmasters in our baseline survey also share the view of teachers that official curriculum should be followed under such circumstances. In fact, 97 percent of headmasters consider ensuring adherence to the curriculum as an important part of their job and 80 percent of them think that they have influence over determining how the teachers deliver the curriculum lessons to students at school. Total enrollment is statistically different across the three arms. Class 9 enrollment in Standard *Utkarsh* schools is smaller than in the other two arms. However, some degree of imbalance is expected and is not expected to pose a challenge to our evaluation.

7 Results

7.1 Student Outcomes

Baseline Achievement We begin by documenting the existing low learning levels of students to provide useful context. In all of the three baseline subjects, the mean student is over 4 Class levels below Class 9; see Appendix Figure A2.[8] Nearly half of all students are evaluated below Class 3. The sample also has substantial heterogeneity–nearly 8 percent of students in Math and English and 18 percent of students in Odia are at grade level. This heterogeneity is both across schools and within schools–in math, the average interquartile ranking of competency (i.e., the mean difference within a school between the 75th and 25th percentiles) is 3.94 grades. Therefore, on average, although nearly half of Class 9 students are below Class 3 competency in a given classroom, remedial instruction is not needed and may even harm the the 8-18 percent of students who are at grade level. Teachers are relatively accurate in their estimates of student proficiency–teachers only overestimate the percent of

[8]Grade level mastery is based on the percent correct of baseline questions from that grade level's curriculum. The exams contained no grade level 1 or 2 material. Students below Class 3 competency are given a competency of Class 2 even thought their learning level might have been Class 1.

their students who are at or above Class 5 by about 6 percentage points [9]

Follow-Up Achievement We now show the main effects of the program on learning. The *Utkarsh* program improved student learning across all four target subjects. Each version of Utkarsh increased students' overall test scores by about 0.11 SD (Table 1, column 1). Over this same period, overall test scores for the control group increased by 0.19 SD. Therefore, Utkarsh improved learning by 59% relative to the status quo, regardless of implementation approach. *Utkarsh's* effects on subject-specific scores are similar to the overall effects: Utkarsh increased English and Math scores by 0.12 SD (columns 2 and 3), Odia scores by 0.09 SD (column 4), and Science scores by 0.10SD (*Standard Utkarsh*) and 0.14 (*Flexible* Utkarsh, column 5). Relative to the control group's rate of test score gains over the same period, the program increased learning in English by 57 percent, Math by 190 percent, and Odia by 43 percent.[10] These results are robust to multiple hypothesis testing corrections (Appendix Table A5). In Appendix Tables A6-A8 we test for heterogeneity by gender, scheduled caste and scheduled tribe, or first generation learner status and find some evidence that *Utkarsh* language instruction was especially beneficial to female students.

Utkarsh increased the grade level of students, but did not bring them to grade level. Recall that at baseline students were on average over 4 grades behind. *Utkarsh* increased the likelihood that students were at least grade 3 or grade 5 in Math and at least grade 5 in Odia (Figure 3). Neither intervention improved the likelihood of Class 8 mastery in any subject. However, the interventions also did not decrease the likelihood of Class 8 mastery, a common concern of remedial programs. One reason that *Utkarsh* may have

[9]These results are similar to teacher-collected data from the program leveling test, which also find a high share of students have low-level of learning (see Appendix Table A3).

[10]We cannot compare science increaess relative to control group learning since student did not take a baaseline science test. In Appendix Table A4 we show that Utkarsh also increased test scores on a subset of 5 questions selected from the PISA in Math and 4 questions selected from the PISA in English, corroborating that the results are not merely due to teaching-to-the-test. Overall PISA scores (combining English and Math questions) improved by 0.07 SD for the Flexible arm and 0.05 SD for the Standard arm, with the former effect being significant at the 5 percent level. Looking at the subject-specific PISA scores, we see larger effects of Standard Utkarsh for English PISA questions, and larger effects of Flexible Utkarsh for Math PISA, although we can only reject the equality of the two effects for the latter, and only at the 10 percent level.

not brought students all the way up to grade level is the relatively short time period and distribution of student learning levels in relation to grade-level competency. See Figure A3. The median student at Grade 3 competency is 0.02 standard deviations above the mean while the median student demonstrating Grade 8 competency is approximately 1.52 standard deviations above the mean. Therefore, to bring students up to grade level would require over a 1.3 standard deviation improvement in test scores, approximately 10 times the observed *Utkarsh* treatment effect.

Students throughout the baseline test score distribution benefited from the program. In Appendix Figure A4, we plot non-parametric test score effects (i.e., kernel-weighted locally smoothed regressions by baseline test score percentile) and find learning gains throughout the baseline test score distribution. When dividing students into terciles based on their baseline test scores, the overall and Odia test score results are positive and statistically equivalent across the three terciles with the largest point values for the students in the lowest tercile. For English, we reject equality across the terciles finding the largest effects for the lowest tercile (0.17SD) and smaller effects for the highest tercile (0.06SD). For Science, we also reject equality across the three terciles, only finding statistically significant effects for the top tercile for *Standard Utkarsh* (0.19SD) and the top two terciles for *Flexible Utkarsh* (effect sizes of 0.20 SD for the middle third and 0.17 SD for the top third); see Appendix Table A9.

Attrition As with any RCT, one concern is attrition at the follow-up generating differential selection into the test. To limit attrition, we attempted to follow-up with all students from the baseline, even those who were not present in school the day of the follow-up visit. As a result, overall attrition was very low, only 6 percent in the control group. Appendix Table A10 estimates differences across study arms in the likelihood that students completed the achievement follow-up. Students in the *Standard Utkarsh* arm were two percentage points more likely to complete the achievement follow-up (column 1), and this is differential by baseline test score with lower scoring baseline students more likely to complete the achievement follow-up in the standard arm (column 2). Because of this differential attrition, we

constructed treatment bounds following Lee (2009). As shown in Appendix Table A11 in all cases both the magnitude and statistical significance are similar to the main effects.

Non-Cognitive Outcomes The *Utkarsh* program could have been encouraging to students because they were being taught at their learning level or discouraging because they were told that they needed remedial attention. We find no evidence of discouragement and limited evidence of encouragement (Appendix Table A12). The treatment does not affect students' self-reported ranking among peers or their estimated Board Exam scores in English, Math, or Odia. Students in the standard arm report marginally higher expected Science scores (1 percentage point over a control group mean of 65 out of 100) and 3 percentage points more likely to desire a bachelor's degree, relative to a control group mean of 51 percent.

7.2 Classroom Practices

Teachers in the treatment arms improved their teaching practices and implemented the *Utkarsh* program (Table 2). The effects on test scores in the previous sub-section were not because teachers were more likely to be present at the start of the school day or teaching during a classroom observation (columns 1 and 2). Instead, the treatments increased a general measure of the quality of classroom teaching by 0.35 (*Flexible*) and 0.39 (*Standard*) standard deviations–classrooms became more active and engaging and more likely to involve interactions between teachers and students.[11] Teachers also implemented specific aspects of *Utkarsh* beyond active pedagogy–on average each school implemented about 81 percent of the fourteen different components (column 4).[12]Headmasters were also no more likely to be in attendance, although point estimates are positive (see Appendix Table A15). Results are also not driven by differential teacher attendance during the unannounced monitoring visit (see Appendix Table A16).

[11]Results for each of the specific items of the teaching practices index are in Appendix Table A13.

[12]These components of *Utkarsh* were common to the two interventions. Results for each of the specific items of the implementation index are in Appendix Table A14. This includes both teacher reported and enumerator observed aspects. Results are similar when limited only to enumerator observed components.

7.3 Curriculum

The previous two subsections showed that student learning increased equally across the two arms as did teaching quality. In this section we address whether the curriculum delivered across the two arms differed–when teachers were given the flexibility to diverge from the standard *Utkarsh*, did they? Table 3 contains these outcomes.

Only about 20 percent of teachers in the *Flexible* arm completed the *Flexible Utkarsh* teaching plan and 15 percent reported that they followed a Flexible plan during the week of observation (columns 1 and 2). Teachers in the Standard arm were to follow the prescribed lessons while those in the *Flexible* arm were instructed to either complete a Flexible plan and follow that or continue to follow the Standard plan. Completing a Flexible plan is neither a necessary nor sufficient condition for engaging in flexibility. Even though almost all teachers in both arms felt like they had autonomy over using *Utkarsh*, a statistically higher portion of *Flexible* teachers agreed (column 3). Teachers in the Flexible arm were also 7 percentage points more likely to say that they could adjust topics or pace if students were struggling with a concept (column 4). Somewhat surprisingly given the emphasis on completing the curriculum, about 77 percent of control teachers also reported that they could do this, the same level as the Standard treatment arm. Therefore, the majority of teachers already feel like they can adjust the course content, but they largely do not, instead relying on teaching a curriculum that is multiple grade levels above their students' learning levels. Thus, the teachers are enacting the wishes of the bureaucrats, even while acknowledging there is limited oversight to enforce that directive.[13]

We further test for the implementation of flexibility by comparing what teachers said they were covering relative to the *Standard Utkarsh* schedule. Table 4 compares teachers in the Flexible arm to those in the Standard arm. About 95 percent of teachers reported

[13]In Appendix Table A17 we examine the characteristics of teachers who "take up" the flexibility. Younger, less experienced teachers are less likely to take up flexibility. There are also some differences by subject and responses to the baseline survey predicting who followed different aspects of the program. In Appendix Table A18 we show that those teachers int he FLexible arm with a wider range of student abilities in the classroom are more likely to respond that they have discretion in the classroom at endline.

doing an *Utkarsh* lesson during the week (column 1). Teachers across both treatments made *Utkarsh* their own–only 43 percent were doing the prescribed Standard lesson for that week (column 2). Relative to the prescribed schedule, Flexible teachers were more likely to deviate widely–82 percent of Standard teachers were within a week of the Standard schedule and Flexible teachers were 9 percentage points less likely to be covering a lesson on that same approximately on time schedule.

Teachers in both arms embraced some level of flexibility. The additional flexibility exerted by teachers in the Flexible arm did not lead to more shirking or differentially change student test scores. Flexibility and discretion are touted as ways to improve worker motivation and decrease burnout. Conversely, this new program could have increased teacher stress and anxiety. We find similar levels of teacher burnout, stress, and anxiety across all three arms. Teachers further did not alter their self reported lesson preparation time or grading time–any *Utkarsh* preparation time was offset by a decrease in curriculum-level preparation time (Appendix Table A19).

7.4 Teachers' Perceptions of the Program, Students, and Themselves

Teachers overwhelmingly believed that they and their students benefited from *Utkarsh* (Table 5, columns 1 and 2). Yet, they lowered their estimates of the percentage of their students who could perform a basic literacy (writing a simple English sentence) or numeracy (three digit minus two digit) operation by 5 percentage points. In addition to lowering their expectations about specific tasks, teachers reduced their expectations about the percent of students who would pass the Class X Board exam by about 5 percentage points from a control group mean of 61 percent (column 6).[14] They also reduced their estimates of how many students would earn a bachelor's degree by about 4 percentage points from the control group estimate of 50 percent (column 7).

[14]Statewide the average annual pass rate from 2012 and 2020 (pre-pandemic) was between 71 and 85 percent.

Despite believing that they and their students benefited from Utkarsh, it did not lead teachers to self-assess themselves as any more effective than teachers at schools similar to theirs perhaps because they assessed students more accurately and realized how little their students knew (Appendix Table A20). The program at most marginally objectively improved teacher content knowledge, improving the percent correct on a math test designed to be at the 4th grade level (Appendix Table A20).[15]

7.5 Board Marks and Longer Run Outcomes

Students in our sample were scheduled to take the high-stakes Class X board exam one year after the *Utkarsh* program ended, in June 2021. Board exams scheduled for June 2021 were canceled due to the Covid-19 pandemic. Instead, students were assigned "Board Marks" based upon a weighted average of their school-based exams from Class 9 (40%) and Class 10 (60%), which were scored by the students' own teachers.[16] Since the board marks were largely determined by teacher perceptions rather than an objective test, changes in teacher perceptions of students in treatment schools could also affect their eventual board exam scores.

The interventions did not change the likelihood that a student received passing Board Marks (Table 6, column 1).[17]In our study schools, the Board pass rate was over 99 percent, similar to the statewide average of 98 percent, which was 13 percentage points higher than a previous historical average. Therefore, these were not typical Board Marks. The Standard treatment lowered the average Marks of students by 0.16 standard deviations. These results are consistent with teachers' diminished perceptions of students as a result of the program that persist well after the program ended, although could also reflect lower grade-level competency. The lower scores in non-*Utkarsh* subjects suggest spillovers from perceptions of the

[15]This test was a measure of teacher competency developed by the World Bank. Teachers were given a homework assignment to grade and were evaluated based upon how many mistakes they caught.

[16]Five percent of students chose to take an actual Board Exam. Their Board Marks are the maximum of the school-based and test-based Marks. We do not know which students sat for a Board Exam.

[17]To maximize power we use the entire sample from our study schools, not only those students who are part of the sample above.

directly targeted subjects (where teachers received information that their students where struggling) onto perceptions of student ability more generally. While these lower scores did not lead to changes in actual pass rates on the exam, primarily due to the special circumstances of the COVID-19 pandemic, they suggest that one key impact of *Utkarsh* is prompting in teachers increased knowledge of how far behind their students are. This interpretation is consistent with the previous responses of teachers generally teaching the curriculum even though students are very far behind, and with reports that teachers could deviate from the curriculum without punishment. It is also consistent with the teacher responses during the period of our evaluation.

As an additional test of longer run outcomes, we conducted a phone survey in November and December 2021.[18] Despite lower average test scores, consistent with the lack of treatment effects on receiving passing Board Marks, the interventions did not affect whether the student was enrolled in school, enrolled in Class 11, or employed (see Table 6, Columns 6-8). Thus, although the treatment caused students to receive lower Board marks, we observe no changes in other long-term outcomes of interest, and specifically find no evidence that students were harmed by the intervention.

8 Cost Effectiveness

As implemented at a 200-school scale, the intervention cost $11.64 per student. As the training and monitoring costs were identical in both arms, both arms were equally cost-effective. The observed cost per student translates into a 0.95 SD overall test score gain per $100 spent. This estimate is comparable to an after-school personalized tutoring intervention known as Mindspark for students primarily in grade 7 and 8 (0.93 SD per $100 at 50-school scale) (Muraldiharan et al., 2019). It is slightly less than the 1.4 SD per $100 at 200-school scale eLearn program which introduced school screens and videos in Pakistan (Beg et al.,

[18]The coverage rate in the phone survey was just 23 percent, not correlated with treatment status (Appendix Table A21).

2022).

9 Discussion

Public sector services in developing countries typically have poor service delivery outcomes, some of which may partly be due to highly regimented service delivery. In this paper we analyze the introduction of a remedial instruction program to help better understand the causes and challenges of improving educational productivity. Our evaluation leads to several important empirical findings. First, we find that at baseline, the mean Class 9 student in our sample is over 4 grade levels behind in math, English, and Odia. Moreover, much of this variation is within classrooms: the typical classroom has a range of student competencies of 3.94 grades, although approximately 10 percent of students are at grade level. With these substantial learning gaps, as well as substantial heterogeneity, it is feared that introducing remedial education programs may crowd out grade-level skills and stall progress for students who are at grade level. While teachers are generally aware of these deficiencies, it is feared that allowing teachers to teach remedial education will be equivalent to lowering standards. It is feared that allowing teachers to teach at the level of the student in their classroom may provide a disincentive for teachers to work hard, and will not help students pass high-stakes tests. Our randomized evaluation empirically evaluates these concerns and finds that they are unfounded. We find that introducing a remedial education known as Utkarsh substantially improved student learning, by 58%, and did not crowd-out grade level competencies (although they also did not change). Thus, one key finding is that in contexts where many students are behind, introducing remedial instruction benefits students without the feared consequences.

Our second empirical finding is that allowing more flexibility in service delivery– and specifically allowing headmasters and teachers agency in what is taught in the classroom- did not meaningfully change its quality; both models were highly effective at delivering the new remedial instruction. Generally the two different implementation models have similarly high

rates of fidelity to the program guidelines, with minimal differences. Part of this response is likely due to the observation that teachers generally adjust lesson timetables and content even without being instructed to do so. While few teachers filled out the teaching plan, meaning that they intended to deviate from the recommended plan at the time of training, during enumerator observations 43 percent of teachers in the *Standard Utkarsh* arm were modifying the timetable. relatively low rates of take-up of the flexibility. These results echo anecdotal evidence that teachers want additional autonomy, and suggest that many teachers will do what they think is best for students even if not directly told. Modifications from the timetables are higher when teachers are explicitly given flexibility, with no end difference in student achievement.

In total, there are several lessons from this study that help provide new evidence on the substantial learning gaps observed. First, many Indian secondary school students are substantially behind grade level. Moreover, our results suggest that teachers are generally aware of these gaps, although there is substantial measurement error in teacher perceptions of student ability. Second, despite a traditional emphasis on curriculum delivery, as well as concerns that teachers may lack adequate competency to modify from prescribed lessons teachers were able to adapt to offering remedial instruction successfully. Moreover, many teachers adapt the their lesson plans and timetable regardless of what is advised, suggesting that teachers generally try and deliver content to students as they see fit. Thus, our results suggest that one reason why service delivery is poor in the public sector is because teachers lack the appropriate materials and direction to allow them to teach at the level of the student.

Second, our study provides guidance on the optimal allocation of authority in the public sector. While offering teachers flexibility did not differentially improve learning, it also did no harm. These results bolster the interpretation that rigid bureaucracies could improve service delivery by modifying rules and potentially explicitly giving teachers increased ability to adapt to local conditions. Despite concerns of coordination challenges of changing the status quo, our results suggest that teachers were able to effectively adapt to a new, more

effective approach in the classroom that benefitted both students and teachers.

Third, we shed light on whether remedial instruction is a wise policy choice at the secondary school level and more generally build upon the scant policy base of what interventions are effective at improving learning at the secondary school level. Despite concerns of crowd-out, or that the secondary school level may be too late to introduce effective learning interventions, we find substantial increases in student achievement: 60% compared to the status quo. Thus, our results suggest that remedial education is a good use of class time in combination with the current curriculum for secondary school students. Our results show that although nearly half of all students are substantially behind the curriculum, it is cost-effective to improve learning at the secondary school level and decrease the substantial heterogeneity observed within a classroom. The *Utkarsh* intervention specifically did not crowd out (nor change) grade-level mastery, a key concern. Teachers also become more accurate in their evaluation of student competencies. One caveat, however, is that teachers may view a trade-off between the curriculum and remedial education, even when one does not exist. While we find that pass rates on the Board exam were unaffected, partially due to near universal pass rates, results are consistent with the interpretation that teachers rated students as less proficient at grade-level competencies. These results have important implications for ensuring that teachers are made aware of how far behind students are, and suggests that more emphasis should be made on disseminating information on how much students are learning so that teachers are not demotivated by remedial programs.

10 Conclusions

In spite of rigid bureaucratic rules to ensure uniform delivery of service and reinforce high standard of performance, public sectors in developing countries are often crippled with poor service deliver and low productivity. We partner with PFA—a non governmental organization—and the state government of Odisha to implement an RCT in 300 secondary schools to evaluate a remedial instructional program (*Utkarsh*) and answer whether allow-

ing public servants (teachers, in this case) more flexibility to tailor their service (teaching remedial lessons) according to the need of their clients (students) improves the productivity (student learning) of the public service.

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. PFA designed the *Utkarsh* program to enable students who were behind grade level in Class 9 to catch up to curriculum-level material on English, math, Odia, and science. 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.

We find that introducing remedial education for part of the school day significantly improved learning, by approximately 60 percent. Moreover, we find that offering teachers flexibility did not differentially improve or harm school productivity—the program improved student learning the same irrespective of teacher was explicitly given more flexibility or not. The positive effect of *Utkarsh* has important policy consequences as the state government is committed to scale-up *Utkarsh* to all 30 districts of Odisha, reaching about 350,000 Class 9 students.

We also do not find any systematic difference in the fidelity of program implementation, assuaging any concern that providing flexibility to teachers might encourage them to shirk. However, we find that teachers generally did not use the flexibility offered, appearing to prefer teaching the lessons as-is.

However, one limitation of our finding is that the learning improvement is still inadequate to cover students' learning gap. After the program, students still suffer grade-level learning gap. This suggests that additional research is required to design interventions that would improve secondary students' learning in similar setting by larger amount.

References

Afridi, Farzana, Vegard Iversen, and M. R. Sharan, "Women Political Leaders, Corruption, and Learning: Evidence from a Large Public Program in India," *Economic Development and Cultural Change*, 2017, *66* (1), 1–30. Publisher: The University of Chicago Press.

Bandiera, Oriana, Michael Carlos Best, Adnan Qadir Khan, and Andrea Prat, "The Allocation of Authority in Organizations: A Field Experiment with Bureaucrats," NBER Working Paper 26733, National Bureau of Economic Research, Cambridge, MA 2020.

Banerjee, Abhijit, Chattopadhyay, Raghabendra, Duflo, Esther, Keniston, Daniel, and Singh, Nina, "Improving Police Performance in Rajasthan, India: Experimental Evidence on Incentives, Managerial Autonomy, and Training," *American Economic Journal: Economic Policy*, 2021, *13* (1), 31–66.

_ , Rukmini Banerji, James Berry, Esther Duflo, Harini Kannan, Shobhini Mukerji, Marc Shotland, and Michael Walton, "From Proof of Concept to Scalable Policies: Challenges and Solutions, with an Application," *Journal of Economic Perspectives*, 2017, *31* (4), 73–102.

Banerjee, Abhijit V, Rukmini Banerji, Esther Duflo, Rachel Glennerster, and Stuti Khemani, "Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India," *American Economic Journal: Economic Policy*, 2010, *2* (1), 1–30.

_ , Shawn Cole, Esther Duflo, and Leigh Linden, "Remedying Education: Evidence from Two Randomized Experiments in India," *Quarterly Journal of Economics*, 2007.

Barrera-Osorio, Felipe and Dhushyanth Raju, "Teacher Performance Pay: Experimental Evidence from Pakistan," *Journal of Public Economics*, 2017, *148*, 75–91.

Beg, Sabrin, Adrienne Lucas, Waqas Halim, and Umar Saif, "Engaging Teachers with Technology Increased Achievement, Bypassing Teachers Did Not," Technical Report w25704, National Bureau of Economic Research, Cambridge, MA 2019.

Beteille, Tara, Elaine Ding, Ezequiel Molina, Adelle Pushparatnam, and Tracy Wilichowski, *Three Principles to Support Teacher Effectiveness During COVID-19*, World Bank, Washington, DC, 2020.

Bjorkman, Martina and Jakob Svensson, "Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monitoring in Uganda," *The Quarterly Journal of Economics*, May 2009, *124* (2), 735–769.

Bloom, Nicholas, Renata Lemos, Raffaella Sadun, and John Van Reenen, "Does Management Matter in Schools?," *The Economic Journal*, 2015, *125* (584), 647–674. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ecoj.12267.

Bold, Tessa, Deon Filmer, Gayle Martin, Ezequiel Molina, Brian Stacy, Christophe Rockmore, Jakob Svensson, and Waly Wane, "Enrollment without Learning: Teacher Effort, Knowledge, and Skill in Primary Schools in Africa," *Journal of Economic Perspectives*, 2017, *31* (4), 185–204.

_ , _ , _ , _ , **Christophe Rockmore, Brian Stacy, Jakob Svensson, and Waly Wane**, "What Do Teachers Know and Do? Does It Matter?," Working Paper 7956, World Bank, Washington, DC 2017.

_ , **Ezequiel Molina, Deon Filmer, and Jakob Svensson**, "The Lost Human Capital: Teacher Knowledge and Student Learning in Africa," 2018. Publisher: CEPR Discussion Paper No. DP12956.

Brown, Christina and Tahir Andrabi, "Inducing Positive Sorting Through Performance Pay: Experimental Evidence from Pakistani Schools," *University of California at Berkeley Working Paper*, 2020.

Chaudhury, Nazmul, Jeffrey Hammer, Michael Kremer, Karthik Muralidharan, and F. Halsey Rogers, "Missing in Action: Teacher and Health Worker Absence in Developing Countries," *Journal of Economic Perspectives*, 2006, *20* (1), 91–116.

Das, Jishnu and Tristan Zajonc, "India Shining and Bharat Drowning: Comparing Two Indian States to the Worldwide Distribution in Mathematics Achievement," *Journal of Development Economics*, 2010, *92* (2), 175–187.

Dessein, Wouter, "Authority and communication in organizations," *The Review of Economic Studies*, 2002, *69* (4), 811–838.

DiNardo, John, Jordan Matsudaira, Justin McCrary, and Lisa Sanbonmatsu, "A Practical Proactive Proposal for Dealing with Attrition: Alternative Approaches and an Empirical Example," *Journal of Labor Economics*, 2021, *39* (S2), S507–S541. Publisher: The University of Chicago Press.

Djaker, Sharnic, Alejandro Ganimian, and Shwetlena Sabarwal, "Primary- and middle-school teachers in South Asia overestimate the performance of their students," *New York University Working Paper*, 2022.

Duflo, Annie, Jessica Kiessel, and Adrienne Lucas, "Experimental Evidence on Alternative Policies to Increase Learning at Scale," Technical Report w27298, National Bureau of Economic Research, Cambridge, MA 2020.

Duflo, Esther, Pascaline Dupas, and Michael Kremer, "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya," *American Economic Review*, 2011, *101* (5), 1739–1774.

_ , _ , **and** _ , "School Governance, Teacher Incentives, and Pupil–Teacher Ratios: Experimental Evidence from Kenyan Primary Schools," *Journal of Public Economics*, 2015, *123*, 92–110.

_ , **Rema Hanna, and Stephen P. Ryan**, "Incentives Work: Getting Teachers to Come to School," *The American Economic Review*, 2012, *102* (4), 1241–1278. Publisher: American Economic Association.

Figlio, David N and Marianne E Page, "School choice and the distributional effects of ability tracking: does separation increase inequality?," *Journal of Urban Economics*, 2002, *51* (3), 497–514.

Gailmard, Sean, "Expertise, subversion, and bureaucratic discretion," *Journal of Law, Economics, and Organization*, 2002, *18* (2), 536–555.

Glewwe, Paul, Michael Kremer, Sylvie Moulin, and Eric Zitzewitz, "Retrospective vs. prospective analyses of school inputs: the case of flip charts in Kenya," *Journal of development Economics*, 2004, *74* (1), 251–268.

_ , **Nauman Ilias, and Michael Kremer**, "Teacher Incentives," *American Economic Journal: Applied Economics*, 2010, *2* (3), 205–227.

Gray-Lobe, Guthrie, Anthony Keats, Michael Kremer, Isaac Mbiti, and Owen W Ozier, "Can education be standardized? Evidence from Kenya," *Evidence from Kenya (September 16, 2022). University of Chicago, Becker Friedman Institute for Economics Working Paper*, 2022, (2022-68).

Jo, Jinhee and Lawrence S Rothenberg, "The importance of bureaucratic hierarchy: conflicting preferences, incomplete control, and policy outcomes," *Economics & Politics*, 2014, *26* (1), 157–183.

Lakshminarayana, Rashmi, Alex Eble, Preetha Bhakta, Chris Frost, Peter Boone, Diana Elbourne, and Vera Mann, "The Support to Rural India's Public Education System (STRIPES) Trial: A Cluster Randomised Controlled Trial of Supplementary Teaching, Learning Material and Material Support," *PLoS ONE*, 2013, *8* (7), e65775.

Lavy, Victor and Analia Schlosser, "Targeted remedial education for underperforming teenagers: Costs and benefits," *Journal of Labor Economics*, 2005, *23* (4), 839–874.

Lee, David S., "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects," *The Review of Economic Studies*, 2009, *76* (3), 1071–1102.

Muralidharan, Karthik, Abhijeet Singh, and Alejandro J. Ganimian, "Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India," *American Economic Review*, April 2019, *109* (4), 1426–1460.

_ **and** _ , "Improving Public Sector Management at Scale? Experimental Evidence on School Governance India," Technical Report w28129, National Bureau of Economic Research, Cambridge, MA 2020.

_ **and Venkatesh Sundararaman**, "Teacher Performance Pay: Experimental Evidence from India," *Journal of Political Economy*, 2011, *119* (1), 39–77.

NITI Aayog, "SDG India Index & Dashboard 2020-21," Technical Report, National Institute of Transforming India, New Dehli 2021.

Pearson, L Carolyn and William Moomaw, "The Relationship Between Teacher Autonomy and Stress, Work Satisfaction, Empowerment, and Professionalism.," *Educational Research Quarterly*, 2005, *29* (1), 38–54. Publisher: ERIC.

Piper, Benjamin, Yasmin Sitabkhan, Jessica Mejia, and Kellie Betts, "Effectiveness of Teachers' Guides in the Global South: Scripting, Learning Outcomes, and Classroom Utilization," Technical Report, RTI Press 2018.

Rasul, Imran and Daniel Rogger, "Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service," *The Economic Journal*, 2018, *128* (608), 413–446. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ecoj.12418.

_ , _ , and Martin J. Williams, "Management and Bureaucratic Effectiveness: Evidence from the Ghanaian Civil Service," Technical Report 8595 2018.

Ree, Joppe De, Karthik Muralidharan, Menno Pradhan, and Halsey Rogers, "Double for Nothing? Experimental Evidence on an Unconditional Teacher Salary Increase in Indonesia," *The Quarterly Journal of Economics*, 2018, *133* (2), 993–1039. Publisher: Oxford University Press.

Skaalvik, Einar M. and Sidsel Skaalvik, "Teacher Self-Efficacy and Perceived Autonomy: Relations with Teacher Engagement, Job Satisfaction, and Emotional Exhaustion," *Psychological Reports*, February 2014, *114* (1), 68–77.

Zhao, Yong, *What works may hurtâSide effects in education*, Teachers College Press, 2018.

Figures

Figure 1: Study Design

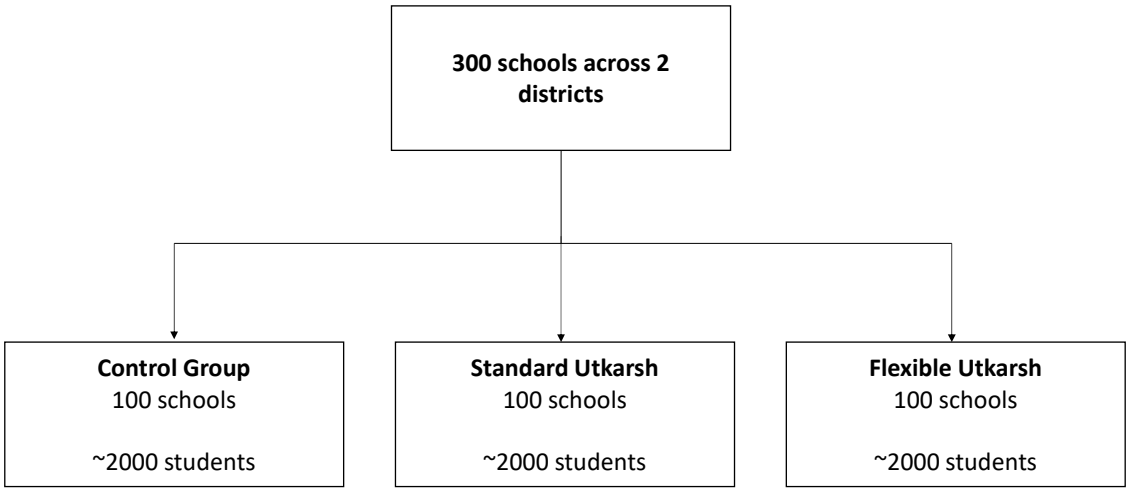


Figure 2: Timeline

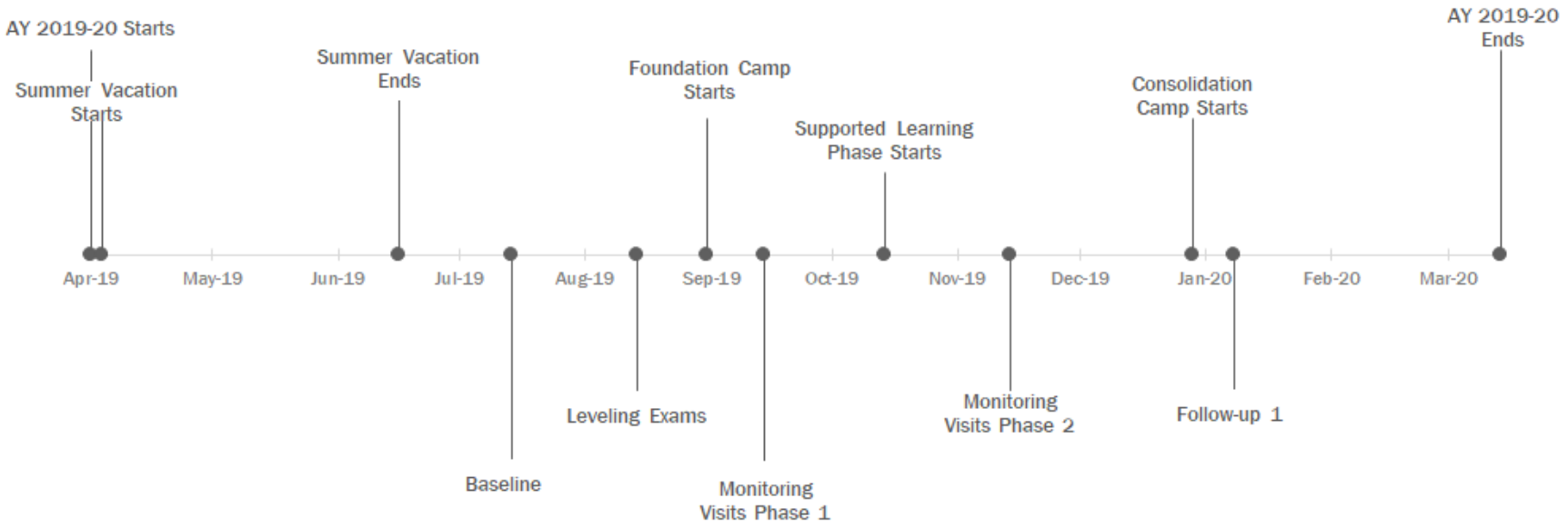
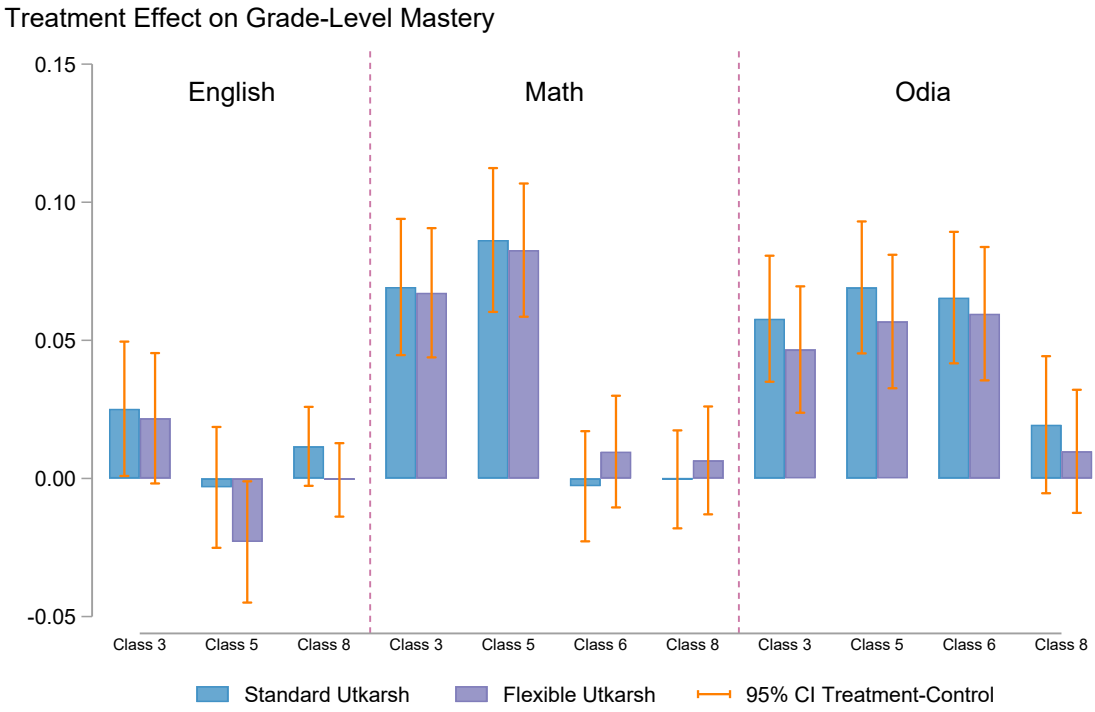
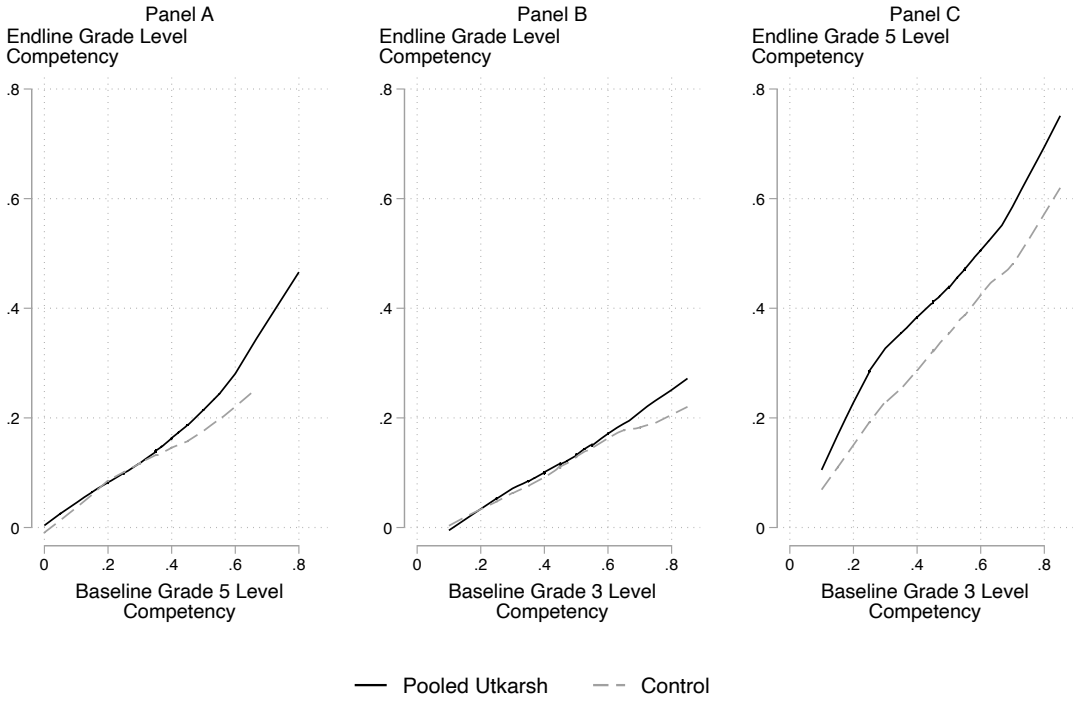


Figure 3: Treatment Effect on Grade-Level Mastery



Notes: This figure is showing the treatment effects on achieving at least respective grade-level mastery in English, math, and Odia. A student is considered to have a specific grade-level mastery in a subject if the student correctly answered at least 75 percent questions that relate to that specific grade's learning level.

Figure 4: Math Competency by Mean Baseline Competency



Notes: This figure is showing school-level distribution of math competency by mean baseline math competency separately for the pooled Utkarsh group and control group. In Panel A, the x-axis is showing the school-level share of students who have grade 5 level competency in math at baseline and the y-axis is showing the school-level share of students who have grade appropriate competency in math at endline. In Panel B, the x-axis is showing the school-level share of grade 3 level competency in math at baseline, while y-axis is the same as Panel A. In Panel C, the x-axis is the same as that of Panel B, while the y-axis is showing the school-level share of students who have grade 5 level competency in math at endline.

Tables

Table 1: Treatment Effects on Students' Test Scores

	Overall	English	Math	Odia	Science
	(1)	(2)	(3)	(4)	(5)
Standard Utkarsh	0.107***	0.118***	0.119***	0.089***	0.104***
	(0.015)	(0.020)	(0.020)	(0.017)	(0.031)
Flexible Utkarsh	0.110***	0.116***	0.123***	0.086***	0.143***
	(0.015)	(0.018)	(0.021)	(0.017)	(0.032)
Observations	5,448	5,448	5,448	5,448	5,448
Control-group change (baseline to endline)	0.19	0.21	0.06	0.21	N/A
Standard Utkarsh=Flexible Utkarsh (p-value)	0.80	0.21	0.85	0.89	0.21

Notes: All regressions include strata, week, and day-of-week fixed effects; student's standardized IRT scores from baseline English, Math, and Odia tests, a dummy for student being female, age of student, and indicator variables for endline interview phase. All columns are showing students' standardized IRT scores from endline tests. Column 1: overall score based on all subject. Columns 2-5: scores in respective subjects. Heteroskedasticity-robust standard errors, clustered at the school level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 2: Impacts on Teacher Behavior

	Teacher in the classroom upon arrival	Teaching	Teaching Practices Index	Implementation Portion
	(1)	(2)	(3)	(4)
Standard Utkarsh	0.032	-0.021	0.385**	0.814***
	(0.052)	(0.030)	(0.159)	(0.010)
Flexible Utkarsh	-0.054	-0.046	0.346**	0.809***
	(0.056)	(0.031)	(0.157)	(0.010)
Observations	299	290	290	299
Control mean	0.84	0.98	0.000	0.00
Standard Utkarsh=Flexible Utkarsh (p-value)	0.11	0.45	0.801	0.73

Notes: All regressions include strata, week, and day-of-week fixed effects. Columns 1-3 are teacher level observations (one teacher observed in each school) and include teacher's age in years and age squared, teacher's years of experience and experience squared, a dummy for teacher being female, a vector of dummy variables for main subject taught by teacher in the baseline, and indicator variables for monitoring visit phase. Outcomes in Columns 1 and 2 are observed by enumerators while teacher was supposed to be teaching in the classroom. Column 3 is constructed from the following list of variables that are observed by enumerators while teacher was supposed to be teaching in the classroom: teaching and learning material visible, at least one student had an opportunity to express their own idea, teacher asked a question to the class, teacher answered students' questions supportively, teacher answered students' questions without showing disrespect, teacher did not ignore students' questions, teacher seemed familiar with content, teacher encourages student, teacher responds to questions satisfactorily, teaching with student interaction. The index is constructed by calculating the proportion of variables that is true for each observation. Column 4 is showing school level implementation fidelity of the Utkarsh program that is constructed from teacher- and school-level variables and includes average teacher age and its square, average teacher experience and its square and share of female teachers. The measure of implementation portion is constructed from the following list of teacher reported variables: leveling assessment was conducted, share of previous 6 days that teacher taught Utkarsh, teacher did Utkarsh worksheet on the day of the survey or the recent most day the teacher taught, implementing the correct phase, whether students attending each phase of Utkarsh met the inclusion criteria of the respective phase, share of all FC lessons that are Utkarsh lesson, teacher followed Utkarsh lesson exactly as instructed in the lesson guides, taught planned lesson during FC, percent of previous 6 days that teacher followed planned Utkarsh during FC, and the following three enumerator observed variables: students using handbooks in class, classroom has a word wall, student desks arranged in small group. Each underlying variable ranges between 0-1 and set to zero for all control group schools. We first take the school-level average of each variable and then take the average over these variables for each school to construct the measure of implementation portion. Heteroskedasticity-robust standard errors, clustered at the school level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table 3: Teacher Flexibility and Discretion

	Filled out flexible Utkarsh teaching plan	Followed flexible Utkarsh teaching plan for this week	Had autonomy in using Utkarsh	Teacher can adjust content if students have difficulty
	(1)	(2)	(3)	(4)
Standard Utkarsh	0.016	0.013	0.944***	0.007
	(0.018)	(0.015)	(0.013)	(0.030)
Flexible Utkarsh	0.201***	0.150***	0.989***	0.071**
	(0.037)	(0.031)	(0.009)	(0.030)
Observations	569	569	834	834
Control group mean	0.00	0.00	0.00	0.77
Standard Utkarsh=Flexible Utkarsh (p-value)	0.00	0.00	0.00	0.03
Source	Directly Observed	Self-reported	Self-reported	Self-reported

Notes: All regressions include strata, week, and day-of-week fixed effects, teacher's age in years and age squared, teacher's years of experience and experience squared, a dummy for teacher being female, and a vector of dummy variables for main subject taught by teacher in the baseline. Columns 1-2, measured at midline, include indicator variables for monitoring visit phase. Column 1 measures whether the teaching plan is filled out. Column 2 is a self-reported measure of whether teachers followed the teaching plan that they filled out for that week. Columns 3-4, measured at endline, includes an indicator variable for early endline visit. Columns 3 is a self-reported measure of teaching having autonomy in using Utkarsh lessons. Column 4 is a self-reported measure of whether teacher can adjust content of the lesson. Heteroskedasticity-robust standard errors, clustered at the school level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table 4: Lesson Topic Relative to Standard Utkarsh

Topic was:	Any Standard Utkarsh lesson	This week's scheduled Standard Utkarsh	Scheduled Standard Utkarsh lesson for this week or an adjacent week
	(1)	(2)	(3)
Flexible Utkarsh	-0.044	0.001	-0.086*
	(0.028)	(0.054)	(0.051)
Observations	289	289	289
Standard Utkarsh mean	0.95	0.43	0.82

Notes: Table is showing results for teachers surveyed at the SLP-phase of midline and restricted to all such teachers who taught English, math, and science Utkarsh lessons We do not include Odia teachers in this table because Odia lessons are reported at a level that is too generalized. All regressions include strata, week, and day-of-week fixed effects, teacher's age in years and age squared, teacher's years of experience and experience squared, a dummy for teacher being female, a vector of dummy variables for main subject taught by teacher in the baseline, and indicator variables for monitoring visit phase. Column 1 is a directly observed snapshot of whether or not the teacher followed any scheduled Standard Utkarsh curriculum during that week of SLP. Column 2 is measuring whether or not the teacher followed that week's Standard Utkarsh during SLP. Column 3 reports whether teacher followed that week's or an adjacent week's Standard Utkarsh curriculum. Heteroskedasticity-robust standard errors, clustered at the school level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table 5: Teachers' Perception of Students and Utkarsh

	Students benefitted from Utkarsh	Teacher benefitted from Utkarsh	*Percent of student who can write a simple English sentence*			*Percent of student who can do a three digit minus two digit substraction*			Teacher Forecasts that ....Percent of Student	
			Teacher estimate	Actual	Teacher estimate-Actual	Teacher estimate	Actual	Teacher estimate-Actual	Will pass the board exam	Will eventually complete a bachelor's degree
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Standard Utkarsh	0.749***	0.868***	-5.679***	3.510**	-9.190***	-5.303***	0.719	-6.023***	-4.463**	-6.297***
	(0.031)	(0.019)	(2.152)	(1.598)	(2.416)	(1.974)	(1.906)	(2.273)	(1.727)	(2.001)
Flexible Utkarsh	0.717***	0.919***	-5.078***	6.346***	-11.424***	-6.259***	1.438	-7.697***	-5.766***	-7.787***
	(0.030)	(0.020)	(1.912)	(1.617)	(2.283)	(1.706)	(2.016)	(2.386)	(1.633)	(1.823)
Observations	834	834	823	823	823	823	823	823	823	823
Control group mean	0.00	0.00	56.76	16.97	39.79	73.35	56.39	16.96	60.64	49.63
Standard Utkarsh=Flexible Utkarsh (p-value)	0.43	0.04	0.76	0.08	0.34	0.61	0.71	0.48	0.46	0.44

Notes: All regressions include teacher's age in years and age squared, years of experience and experience squared, a dummy for being female, a vector of dummy variables for main subject taught in the baseline, and indicator variables for early endline visit. Columns 4 and 7 are actual school-level averages. Heteroskedasticity-robust standard errors, clustered at the school level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

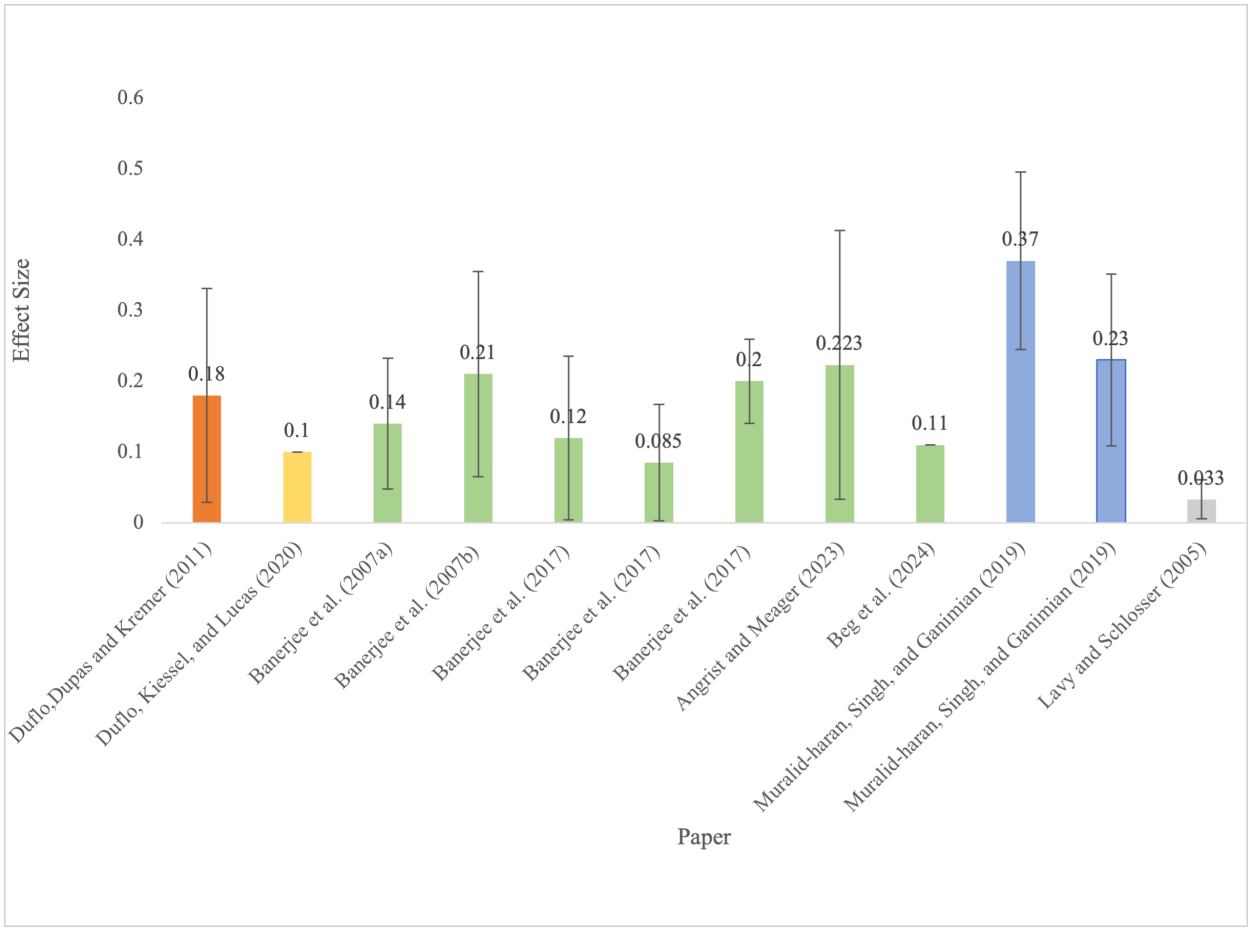
Table 6: Effects on Board Exam Marks and Status After Board Exam

	Passed	*Test Score*			*Grade*	*Status After Board Exam*		
		Total	Utkarsh total	Non-Utkarsh total	B or above	Ernolled in school	Enrolled in class 11	Employed
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Standard Utkarsh	-0.003	-0.157***	-0.134***	-0.192***	-0.056***	0.018	0.013	0.000
	(0.003)	(0.048)	(0.047)	(0.054)	(0.019)	(0.020)	(0.020)	(0.012)
Flexible Utkarsh	-0.004	-0.081	-0.081*	-0.076	-0.021	-0.009	-0.020	-0.004
	(0.003)	(0.049)	(0.048)	(0.056)	(0.020)	(0.020)	(0.021)	(0.011)
Observations	18,551	18,551	18,551	18,551	18,551	1,255	1,255	1,255
Control group mean	0.99	0.00	0.00	0.00	0.39	0.89	0.88	0.04
Standard Utkarsh=Flexible Utkarsh (p-value)	0.88	0.14	0.30	0.04	0.11	0.15	0.11	0.66
Raw control mean		0.56	0.55	0.57				
Raw control SD		0.14	0.15	0.15				

Notes: All regressions include strata fixed effects; standardized IRT scores from baseline English, Math, and Odia tests, a dummy for student being female, and age of student. Columns 1-5: outcomes are from administrative data on board exam results. Column 1 indicates whether the student passed the exam. Columns 2-4 are total scores in all subjects, Utkarsh subjects, and non-Utkarsh subjects, respectively. Colum 5 is indicating whether the student achieved a grade of B or above. Columns 6-8 are from researchers' follow-up survey conducted after the board exam results were published. Column 6 measures whether the student is enrolled in school at all, while Column 7 measures whether the student is enrolled in class 11. Column 8 measures whether the student is employed. Heteroskedasticity-robust standard errors, clustered at the school level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Appendix A Additional Figures

Figure A1: Previous Literature on Remedial Education



Notes: This figure shows the effect size for recent papers that evaluate remedial education programs. Orange indicates that the students in the sample are in Grade 1. Yellow indicates that the students are in lower Primary (Grade 1-3). Geren indicates that the students are in upper primary (grade 3-5). Blue indicates that the students are in junior high (grades 6-8). Grey indicates the students are in high school (grades 10-12).

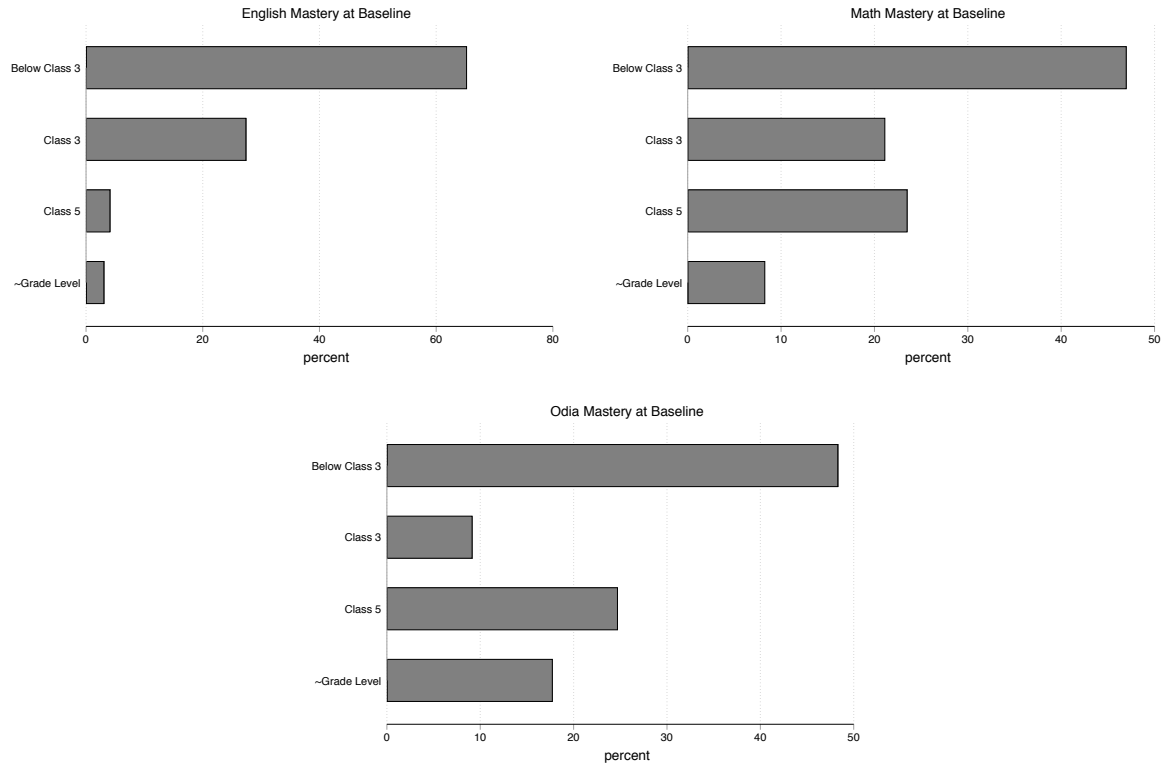
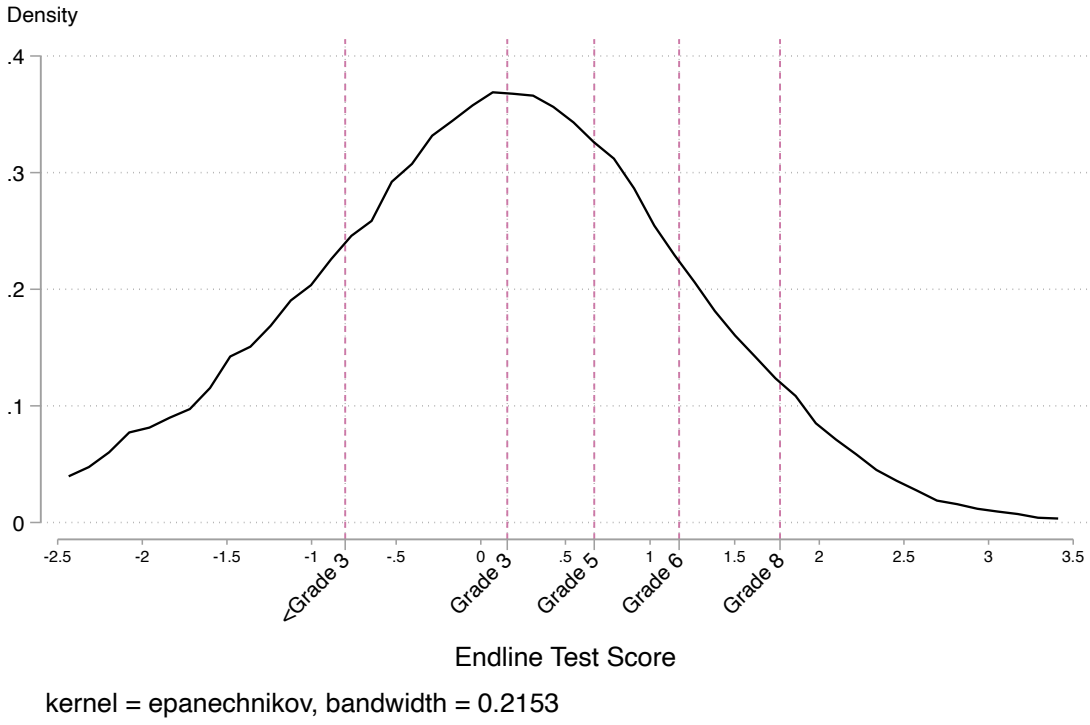


Figure A2: Baseline measures of student ability, by grade level competency

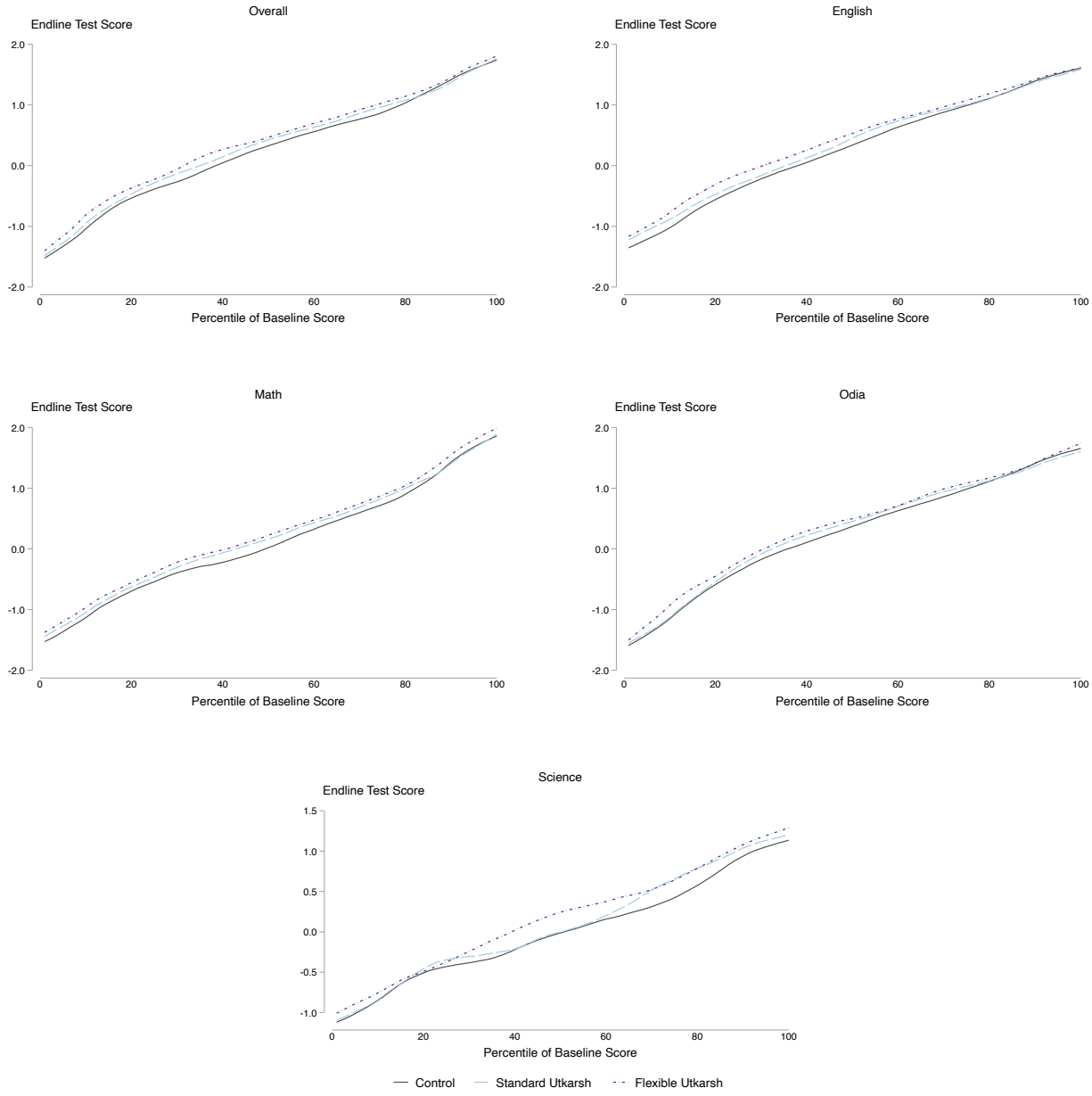
Notes: The figure shows unweighted local mean results of students, categorized by their learning level as described in the text. Results are presented by English, Math, and Odia (local language). The Science test was not administered at baseline.

Figure A3: Control Group Math Learning



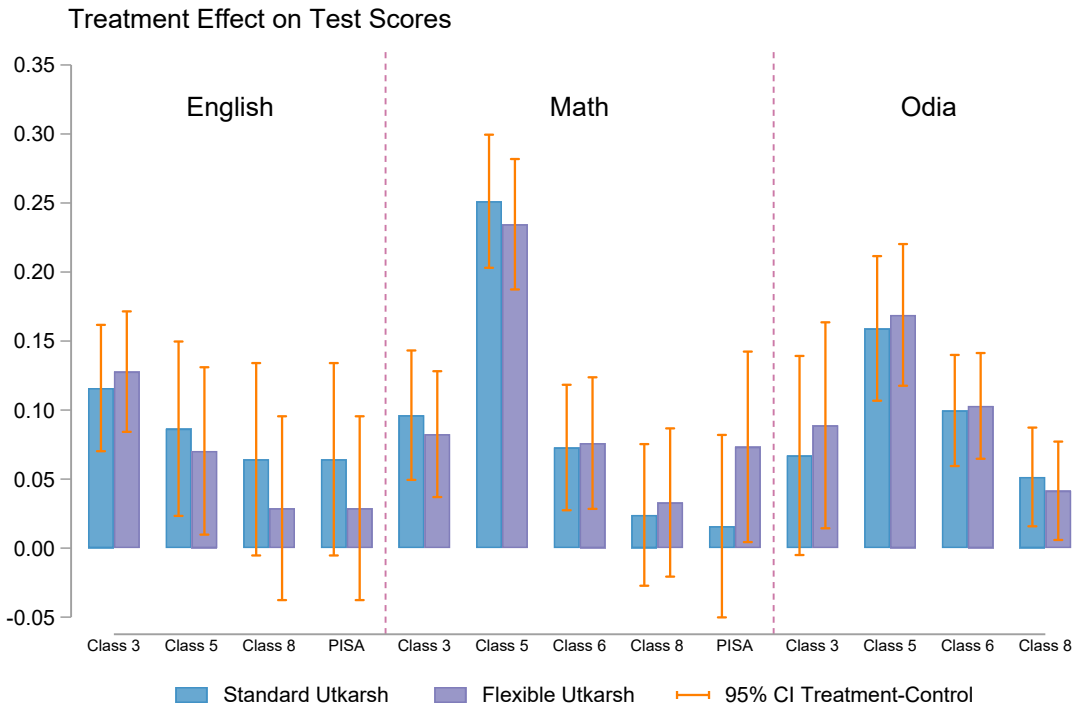
Notes: This figure is showing the distribution of learning of math by control group students. The vertical lines are showing learning levels that correspond to respective grade-level learning.

Figure A4: Non-parametric Distribution of Test Scores by Study Arms



Notes: The figure shows kernel-weighted local mean smoothed distributions of endline test scores over baseline test score percentiles by study arms. Since there was no science test at baseline, we use the average of all subjects at baseline as a proxy for science's baseline. Test scores of standard Utkarsh students at low- and mid-baseline-score-percentiles are higher than the control group in all subjects across. Flexible Utkarsh students' endline scores in all subjects are higher than the control group across the entire distribution of the baseline test scores.

Figure A5: Subject-wise Test Scores in Grade-Level Questions



Notes: This figure is showing the treatment effects on test scores in grade-level questions in English, math, and Odia.

Appendix B Additional Tables

Table A1: Balance Table (Student and School)

	Control	Standard Utkarsh	Flexible Utkarsh	p-value from test of equality (1)=(2)=(3)
Panel A: Student-Level Variables	(1)	(2)	(3)	(4)
Female (=1)	0.50	0.49	0.51	0.61
Age (in years)	13.16	13.15	13.15	0.88
Scheduled caste, scheduled tribe, or other backward caste (=1)	0.60	0.59	0.64	0.06*
No parent can read and write (=1)	0.14	0.16	0.15	0.64
Participated in Utthan (=1)	0.88	0.89	0.88	0.40
Takes private tuition (=1)	0.73	0.71	0.73	0.71
Baseline test scores				
English	0.01	-0.05	0.05	0.09*
Math	-0.01	-0.02	0.03	0.62
Odia	-0.01	-0.02	0.03	0.52
Baseline competency grade				
English	2.65	2.53	2.58	0.04**
Math	3.45	3.37	3.45	0.49
Odia	3.93	3.84	3.93	0.54
Observations	1,949	1,876	1,931	
Panel B: Headmaster- and School-Level Variables	(1)	(2)	(3)	(4)
Female headmaster (=1)	0.26	0.20	0.19	0.46
Age of the headmaster (in years)	52.45	52.44	52.44	1.00
Experience in current position (years)	5.53	5.83	4.70	0.43
Headmaster thinks/considers				
Teacher should follow curriculum (=1)	0.74	0.76	0.80	0.56
Ensuring adherence to curriculum as important job component (=1)	0.95	0.98	0.96	0.45
They have influence on determining delivering curriculum lessons (=1)	0.84	0.79	0.77	0.41
Head of school is the headmaster or headmaster-in-charge (=1)	0.95	0.92	0.92	0.59
Sanctioned class 9 teacher posts in the school	7.55	7.83	7.38	0.31
Number of teacher posts filled	5.19	4.98	5.16	0.57
Total enrollment in class 9	72.38	62.51	72.87	0.02**
Observations	99	100	100	

Notes: Table is showing reported characteristics of the respondents during the baseline survey. Panel A is showing balance of student-level variables. Panel B reports balance of headmaster- and school-level variables. We include strata fixed effects and standard errors are clustered at the school level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels.

Table A2: Balance Table (Teacher)

Panel A: Teachers Surveyed During Baseline	Control (1)	Standard Utkarsh (2)	Flexible Utkarsh (3)	p-value from test of equality (1)=(2)=(3) (4)
Female (=1)	0.48	0.47	0.49	0.96
Age of the teacher (in years)	41.39	41.71	40.87	0.79
Has a teaching certificate (=1)	0.35	0.35	0.35	0.95
Teaching experience (years)	15.77	16.64	15.16	0.48
Teaches an Utkarsh subject (=1)	1.00	1.00	1.00	0.42
Days absent from work	0.33	0.38	0.33	0.73
Works in another school(=1)	0.02	0.02	0.01	0.73
Time spent in preparing for lesson (hours/week)	12.13	12.72	12.78	0.68
Time spent grading (hours/week)	8.46	8.03	8.40	0.64
According to the teacher, percent of student who				
Will pass board exam in first try	63.90	61.86	65.19	0.50
Can write a simple English sentence	55.72	54.79	57.58	0.48
Can do a three digits sum	71.08	73.79	73.02	0.27
Select teaching materials, methods, strategies (=1)	0.95	0.96	0.93	0.38
Allowed to modify course timetable (=1)	0.81	0.80	0.77	0.47
Others do not select evaluation activities (=1)	0.67	0.66	0.65	0.96
Feel pressure to complete curriculum during school year (=1)	0.29	0.28	0.30	0.90
Burnout index	-0.05	-0.15	-0.10	0.64
Autonomy index	0.00	-0.20	-0.16	0.10*
Teacher feels				
Curriculum should be followed even if students have lower learning level (=1)	0.60	0.55	0.62	0.31
That if students' not being ready for board exam would be teacher's own fault (=1)	0.48	0.52	0.47	0.70
Valued and appreciated (=1)	0.69	0.70	0.76	0.09*
Satisfied with job (=1)	0.85	0.82	0.84	0.81
That their opinion seems to count (=1)	0.89	0.92	0.91	0.59
That they have the materials and equipment to teach effectively (=1)	0.63	0.60	0.71	0.08*
Similarly or more effective compared to colleagues (=1)	0.94	0.95	0.96	0.51
Observations	209	189	207	
Panel B: Teachers Surveyed During Monitoring Visit	(1)	(2)	(3)	(4)
Female (=1)	0.47	0.42	0.47	0.55
Age of the teacher (in years)	41.83	42.25	41.60	0.95
Has a teaching certificate (=1)	0.35	0.35	0.35	0.94
Teaching experience (years)	15.77	16.64	15.16	0.48
Teaches an Utkarsh subject (=1)	1.00	1.00	1.00	0.49
Observations	309	306	308	
Panel C: Teachers Surveyed During Endline	(1)	(2)	(3)	(4)
Female (=1)	0.47	0.43	0.47	0.71
Age of the teacher (in years)	41.90	41.99	41.68	0.99
Has a teaching certificate (=1)	0.36	0.35	0.36	0.93
Teaching experience (years)	15.89	16.58	15.25	0.49
Teaches an Utkarsh subject (=1)	1.00	1.00	1.00	N/A
Observations	291	290	289	

Notes: Table is showing reported baseline characteristics of the analysis sample teachers. If any variable is missing from baseline, then we replace that from the midline or endline survey. Panel A is restricted to teachers who were surveyed during the baseline. Panel B reports the characteristics of teachers who were surveyed during the midline survey. Panel C is restricted to teachers who were surveyed during the endline. We include strata fixed effects and standard errors are clustered at the school level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels.

Table A3: Percent of Student in Each Level

Student Level	*Baseline*							
	English		Math		Odia		Science	
	Standard	Flexible	Standard	Flexible	Standard	Flexible	Standard	Flexible
Inception	30.11	31.86	42.45	43.63	27.62	32.46	50.17	50.78
Class 3	24.39	22.62	36.35	37.69	6.52	7.67	26.84	27.05
Class 5	25.51	28.44	6.90	6.70	18.31	18.51	10.87	10.93
Class 8	10.84	8.40	5.12	3.27	36.42	31.39	3.88	2.47
Absent	9.15	8.68	9.18	8.71	11.13	9.97	8.23	8.77
Total Number of Students	7,287	6,569	7,287	6,569	7,287	6,569	7,287	6,569
Student Level	*Endline*							
	English		Math		Odia		Science	
	Standard	Flexible	Standard	Flexible	Standard	Flexible	Standard	Flexible
Inception	15.61	18.50	23.41	23.31	13.69	17.05	37.63	36.76
Class 3	19.25	20.76	34.42	36.40	5.74	5.82	18.07	17.66
Class 5	25.71	25.74	10.83	10.60	12.10	12.28	14.81	16.47
Class 8	24.67	20.76	16.02	14.01	51.67	48.58	14.55	14.98
Absent	14.76	14.23	15.32	15.69	16.80	16.27	14.94	14.13
Total Number of Students	7,284	6,569	7,284	6,569	7,284	6,569	7,284	6,569

Notes: The table shows the share of students at each subject-specific learning level in the baseline and endline of the test that the program administered (*not the test that the authors of this article developed*).

Table A4: Treatment Effects on Students' PISA Test Scores

	Overall	English	Math
	(1)	(2)	(3)
Standard Utkarsh	0.047	0.064*	0.016
	(0.032)	(0.035)	(0.034)
Flexible Utkarsh	0.071**	0.029	0.073**
	(0.032)	(0.034)	(0.035)
Observations	5,448	5,448	5,448
Control-group change (baseline to endline)	0.22	0.15	0.19
Standard Utkarsh=Flexible Utkarsh (p-value)	0.50	0.33	0.09

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 student being female, age of student, and indicator variables for endline interview phase. All columns are students' standardized IRT scores in PISA questions of endline tests. PISA questions were only included in English and math tests. Column 1: overall score based on English and math PISA questions. Columns 2-3: scores in respective subjects' PISA questions. Heteroskedasticity-robust standard errors, clustered at the school level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table A5: Multiple Hypothesis Test - Treatment Effects on Test Scores

	Overall	English	Math	Odia	Science
	(1)	(2)	(3)	(4)	(5)
Standard Utkarsh	0.107***	0.118***	0.119***	0.089***	0.104***
	(0.015)	(0.020)	(0.020)	(0.017)	(0.031)
Flexible Utkarsh	0.110***	0.116***	0.123***	0.086***	0.143***
	(0.015)	(0.018)	(0.021)	(0.017)	(0.032)
Observations	5,448	5,448	5,448	5,448	5,448
Standard Utkarsh=Control (naive p-value)	0.000	0.000	0.000	0.000	0.001
Standard Utkarsh=Control (MHT p-value)	0.000	0.000	0.000	0.000	0.001
Flexible Utkarsh=Control (naive p-value)	0.000	0.000	0.000	0.000	0.000
Flexible Utkarsh=Control (MHT p-value)	0.000	0.000	0.000	0.000	0.000
Flexible Utkarsh=Standard Utkarsh (naive p-value)	0.805	0.904	0.851	0.885	0.210
Flexible Utkarsh=Standard Utkarsh (MHT p-value)	0.997	0.997	0.997	0.997	0.615

Notes: All regressions include strata, week, and day-of-week fixed effects; student's standardized IRT scores from baseline English, Math, and Odia tests, a dummy for student being female, age of student, and indicator variables for endline interview phase. All columns are showing students' standardized IRT scores from endline tests. Column 1: overall score based on all subjects. Columns 2-5: scores in respective subjects. Heteroskedasticity-robust standard errors, clustered at the school level, in parentheses.$^{***}p < 0.01$,$^{**}p < 0.05$,$^{*}p < 0.10$

Table A6: Treatment Effect Heterogeneity by Student Gender

	Overall	English	Math	Odia	Science
	(1)	(2)	(3)	(4)	(5)
Standard Utkarsh	0.090***	0.091***	0.133***	0.064***	0.095**
	(0.019)	(0.024)	(0.028)	(0.022)	(0.039)
Standard Utkarsh x Female	0.033	0.055**	-0.026	0.050*	0.018
	(0.021)	(0.028)	(0.033)	(0.029)	(0.050)
Flexible Utkarsh	0.105***	0.114***	0.145***	0.061***	0.126***
	(0.019)	(0.023)	(0.028)	(0.021)	(0.042)
Flexible Utkarsh x Female	0.011	0.006	-0.043	0.050*	0.033
	(0.022)	(0.031)	(0.035)	(0.026)	(0.052)
Female	-0.017	-0.031	-0.086***	0.071***	-0.089**
	(0.015)	(0.022)	(0.023)	(0.020)	(0.034)
Observations	5,448	5,448	5,448	5,448	5,448
Control-group change (baseline to endline)	0.19	0.21	0.06	0.21	N/A
Standard Utkarsh+Standard Utkarsh x Female=0 (p-value)	0.00	0.00	0.00	0.00	0.00
Flexible Utkarsh+Flexible Utkarsh x Female=0 (p-value)	0.00	0.00	0.00	0.00	0.00

Notes: All regressions include strata, week, and day-of-week fixed effects; student's standardized IRT scores from baseline English, Math, and Odia tests, a dummy for student being female, age of student, and indicator variables for endline interview phase. All columns are showing students' standardized IRT scores from endline tests. Column 1: overall score based on all subject. Columns 2-5: scores in respective subjects. Heteroskedasticity-robust standard errors, clustered at the school level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table A7: Treatment Effect Heterogeneity by Student Caste

	Overall	English	Math	Odia	Science
	(1)	(2)	(3)	(4)	(5)
Standard Utkarsh	0.100***	0.105***	0.122***	0.075***	0.131***
	(0.019)	(0.026)	(0.026)	(0.024)	(0.038)
Standard Utkarsh x SC/ST/OBC	0.012	0.024	-0.005	0.025	-0.050
	(0.021)	(0.030)	(0.030)	(0.028)	(0.048)
Flexible Utkarsh	0.100***	0.085***	0.142***	0.076***	0.122***
	(0.019)	(0.025)	(0.028)	(0.023)	(0.042)
Flexible Utkarsh x SC/ST/OBC	0.017	0.056*	-0.031	0.017	0.033
	(0.022)	(0.029)	(0.033)	(0.028)	(0.051)
SC/ST/OBC (=1)	-0.014	-0.035	-0.015	-0.002	0.015
	(0.015)	(0.021)	(0.023)	(0.019)	(0.034)
Observations	5,448	5,448	5,448	5,448	5,448
Control-group change (baseline to endline)	0.19	0.21	0.06	0.21	N/A
Standard Utkarsh+Standard Utkarsh x SC/ST/OBC=0 (p-value)	0.00	0.00	0.00	0.00	0.04
Flexible Utkarsh+Flexible Utkarsh x SC/ST/OBC=0 (p-value)	0.00	0.00	0.00	0.00	0.00

Notes: All regressions include strata, week, and day-of-week fixed effects; students' standardized IRT scores from baseline English, Math, and Odia tests, a dummy for students belonging to ST/SC/OBC, a dummy for student being female, age of student, and indicator variables for endline interview phase. All columns are showing students' standardized IRT scores from endline tests. Column 1: overall score based on all subjects. Columns 2-5: scores in respective subjects. Heteroskedasticity-robust standard errors, clustered at the school level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table A8: Treatment Effect Heterogeneity by Student's First Generation Learner Status

	Overall	English	Math	Odia	Science
	(1)	(2)	(3)	(4)	(5)
Standard Utkarsh	0.109***	0.123***	0.117***	0.085***	0.116***
	(0.015)	(0.020)	(0.020)	(0.018)	(0.033)
Standard Utkarsh x First Generation Learner	-0.013	-0.031	0.019	0.029	-0.088
	(0.037)	(0.047)	(0.048)	(0.042)	(0.075)
Flexible Utkarsh	0.111***	0.119***	0.122***	0.076***	0.163***
	(0.015)	(0.018)	(0.021)	(0.017)	(0.034)
Standard Utkarsh x First Generation Learner	0.001	-0.015	0.011	0.076*	-0.146*
	(0.034)	(0.044)	(0.051)	(0.041)	(0.077)
First Generation Learner (=1)	-0.046*	-0.054	-0.068**	-0.090***	0.104*
	(0.025)	(0.034)	(0.035)	(0.031)	(0.057)
Observations	5,448	5,448	5,448	5,448	5,448
Control-group change (baseline to endline)	0.19	0.21	0.06	0.21	N/A
Standard Utkarsh + Standard Utkarsh x First Generation Learner=0 (p-value)	0.01	0.05	0.01	0.01	0.69
Flexible Utkarsh+Flexible Utkarsh x First Generation Learner=0 (p-value)	0.00	0.02	0.01	0.00	0.81

Notes: All regressions include strata, week, and day-of-week fixed effects; students' standardized IRT scores from baseline English, Math, and Odia tests, a dummy for students being a first generation learner, a dummy for students being female, age of student, and indicator variables for endline interview phase. All columns are showing students' standardized IRT scores from endline tests. Column 1: overall score based on all subjects. Columns 2-5: scores in respective subjects. Heteroskedasticity-robust standard errors, clustered at the school level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table A9: Treatment Effect Heterogeneity by Baseline Test Scores

	Overall	English	Math	Odia	Science
	(1)	(2)	(3)	(4)	(5)
Standard Utkarsh	0.125***	0.166***	0.111***	0.100***	0.090*
	(0.028)	(0.033)	(0.030)	(0.032)	(0.054)
Standard Utkarsh x BL Test Score Middle Third	-0.021	-0.042	0.055	0.001	-0.063
	(0.028)	(0.038)	(0.036)	(0.036)	(0.064)
Standard Utkarsh x BL Test Score Top Third	-0.037	-0.109***	-0.033	-0.044	0.104
	(0.030)	(0.037)	(0.040)	(0.040)	(0.071)
Flexible Utkarsh	0.129***	0.178***	0.126***	0.122***	0.060
	(0.027)	(0.032)	(0.035)	(0.031)	(0.051)
Flexible Utkarsh x BL Test Score Middle Third	-0.030	-0.063*	-0.016	-0.051	0.127**
	(0.027)	(0.037)	(0.038)	(0.038)	(0.062)
Flexible Utkarsh x BL Test Score Top Third	-0.028	-0.122***	0.016	-0.056	0.116*
	(0.031)	(0.035)	(0.044)	(0.038)	(0.064)
BL Test Score Middle Third	0.082***	0.170***	-0.082***	0.155***	-0.090*
	(0.025)	(0.031)	(0.032)	(0.031)	(0.051)
BL Test Score Top Third	0.087**	0.238***	0.048	0.109***	-0.017
	(0.036)	(0.039)	(0.041)	(0.042)	(0.069)
Observations	5,448	5,448	5,448	5,448	5,448
Control-group change (baseline to endline)	0.19	0.21	0.06	0.21	N/A
Standard Utkarsh+Standard Standard x Middle Third (p-value)	0.00	0.00	0.00	0.00	0.53
Standard Utkarsh+Standard Standard x Top Third (p-value)	0.00	0.01	0.01	0.03	0.00
Flexible Utkarsh+Flexible Standard x Middle Third (p-value)	0.00	0.00	0.00	0.00	0.00
Flexible Utkarsh+Flexible Utkarsh x Top Third (p-value)	0.00	0.01	0.00	0.01	0.00

Notes: All regressions include strata, week, and day-of-week fixed effects; student's standardized IRT scores from baseline English, math, and Odia tests, a dummy for student being female, age of student, and indicator variables for endline interview phase. All columns are showing students' standardized IRT scores from endline tests. Column 1: overall score based on all enline subject tests. The regression includes an indicator for tercile of overall baseline test score and an interaction of treatment status with that indicator. Since there is no science baseline, overall baseline test score is calculated based on baseline English, math, and Odia tests. Columns 2-5: endline scores in respective subjects. Columns 2-4: Include an indicator for tercile of respective subject's baseline test score and an interaction of treatment status with that indicator. Column 5: Includes an indicator for tercile of overall baseline test score and an interaction of treatment status with that indicator. Heteroskedasticity-robust standard errors, clustered at the school level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table A10: Attrition by Treatment Arms

	Student's test scores collected at endline (1)	Student's test scores collected at endline (2)
Standard Utkarsh	0.02**	0.02**
	(0.01)	(0.01)
Flexible Utkarsh	0.01	0.01
	(0.01)	(0.01)
Standard Utkarsh x Overall baseline test score		-0.02*
		(0.01)
Flexible Utkarsh x Overall baseline test score		-0.00
		(0.01)
Overall baseline test score		0.07
		(0.06)
Observations	5,756	5,756
Control Mean	0.94	0.94
Standard Utkarsh=Flexible Utkarsh (p-value)	0.27	0.28
Adjusted R-squared	0.11	0.11

Notes: All regressions include strata fixed effects, standardized IRT scores from baseline English, math, and Odia tests, a dummy for student being female, and age of the student. In all regressions, the outcome variable is equal to 1 if a student's all four test scores are collected during the endline survey. Heteroskedasticity-robust standard errors, clustered at the school level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table A11: Lee Bounds

	Overall Test Score		English Test Score		Math Test Score		Odia Test Score		Science Test Score	
	Lower Bound	Upper Bound	Lower Bound	Upper Bound	Lower Bound	Upper Bound	Lower Bound	Upper Bound	Lower Bound	Upper Bound
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Standard Utkarsh	0.103***	0.124***	0.110***	0.142***	0.106***	0.133***	0.079***	0.106***	0.075**	0.134***
	(0.015)	(0.015)	(0.020)	(0.019)	(0.020)	(0.019)	(0.017)	(0.017)	(0.030)	(0.031)
Flexible Utkarsh	0.106***	0.123***	0.110***	0.131***	0.113***	0.130***	0.081***	0.098***	0.125***	0.167***
	(0.015)	(0.014)	(0.018)	(0.018)	(0.021)	(0.021)	(0.017)	(0.017)	(0.031)	(0.032)
Observations	5,382	5,383	5,382	5,383	5,382	5,383	5,382	5,383	5,382	5,383
Control-group change (baseline to endline)	0.19	0.19	0.21	0.21	0.06	0.06	0.21	0.21	N/A	N/A
Standard Utkarsh = Flexible Utkarsh (p-value)	0.87	0.98	1.00	0.49	0.74	0.90	0.87	0.60	0.09	0.28

Notes: Table is showing results for teachers surveyed at the SLP-phase of midline and restricted to all such teachers who taught English, math, Odia, and science Utkarsh lessons All regressions include strata, week, and day-of-week fixed effects, teacher's age in years and age squared, teacher's years of experience and experience squared, a dummy for teacher being female, a vector of dummy variables for main subject taught by teacher in the baseline, and indicator variables for monitoring visit phase. Column 1 is a directly observed snapshot of whether or not the teacher followed any scheduled Standard Utkarsh curriculum during that week of SLP. Column 2 is measuring whether or not the teacher followed that week's Standard Utkarsh during SLP. Column 3 reports whether teacher followed that week's or an adjacent week's Standard Utkarsh curriculum. Heteroskedasticity-robust standard errors, clustered at the school level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table A12: Student's Self-Assessment and Expected Educational Outcomes

	Student Attendance	Self-assessed Rank among Peers (1-10)	Expected Class 10 Board Exam Score (Out of 100)					Hopes to Achieve Bachelor's degree or above
			Overall	English	Math	Odia	Science	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Standard Utkarsh	0.017	-0.067	0.496	0.093	0.314	0.576	0.957*	0.030*
	(0.022)	(0.047)	(0.469)	(0.567)	(0.592)	(0.553)	(0.537)	(0.017)
Flexible Utkarsh	0.009	-0.050	0.701	0.525	0.757	0.710	0.806	0.007
	(0.022)	(0.055)	(0.459)	(0.582)	(0.576)	(0.521)	(0.522)	(0.017)
Observations	5,710	5,397	5,397	5,397	5,397	5,397	5,397	5,397
Control mean	0.66	4.06	66.20	62.67	66.77	70.48	64.86	0.51
Standard Utkarsh=Flexible Utkarsh (p-value)	0.70	0.76	0.69	0.47	0.46	0.82	0.80	0.17

Notes: All regressions include strata, survey week, and survey day-of-week fixed effects; standardized IRT scores from baseline English, math, and Odia tests, a dummy for student being female, and age of the student. Column 1: Outcome variable measures whether the student was present at school during monitoring visit. Columns 2-8: Outcomes are measured during endline survey. Regressions include outcome variables measured at baseline. Column 2: self-assessed rank among peers where 1 indicates the best rank and 10 indicates the worst student. Outcome variables in Columns 3-7: self-reported expected scores (out of 100) in the board exam at the end of class 10. Column 8: whether the student hopes to achieve a Bachelor's degree or above. Heteroskedasticity-robust standard errors, clustered at the school level, in parentheses.q *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table A13: Components of Teaching Practices Index

	Teaching and learning material visible	At least one student had an opportunity to express their own idea	Teacher asked a question to the class	Teacher answered students' questions supportively	Teacher answered students' questions without showing disrespect	Teacher did not ignore students' questions	Teacher seemed familiar with content	Teacher encourages student	Teacher responds to questions satisfactorily	Teaching with student interaction
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Standard Utkarsh	0.012	0.181**	-0.032	0.075	0.008	0.006	-0.009	0.162**	0.130*	0.146**
	(0.012)	(0.073)	(0.049)	(0.076)	(0.027)	(0.028)	(0.029)	(0.075)	(0.078)	(0.071)
Flexible Utkarsh	0.013	0.186**	-0.043	0.051	0.004	-0.005	-0.037	0.149**	0.159**	0.134*
	(0.013)	(0.078)	(0.045)	(0.077)	(0.028)	(0.026)	(0.031)	(0.075)	(0.076)	(0.073)
Observations	290	290	290	290	290	290	290	290	290	290
Control mean	0.99	0.49	0.89	0.53	0.96	0.97	0.98	0.29	0.46	0.57
Standard Utkarsh=Flexible Utkarsh (p-value)	0.91	0.94	0.81	0.76	0.86	0.61	0.40	0.87	0.70	0.85

Notes: All outcomes are teacher level observations (one teacher observed in each school) during midline and are observed by enumerators while teacher was supposed to be teaching in the classroom. All regressions include teacher's age in years and age squared, teacher's years of experience and experience squared, a dummy for teacher being female, a vector of dummy variables for main subject taught by teacher in the baseline, and indicator variables for monitoring visit phase. Heteroskedasticity-robust standard errors, clustered at the school level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table A14: Implementation of Utkarsh

	Outcomes	Standard Utkarsh	Flexible Utkarsh	N	Standard Utkarsh= Flexible Utkarsh (p-value)
		(1)	(2)	(3)	(4)
(1)	Conducted levelling assessment	0.945***	0.956***	299	0.60
		(0.015)	(0.015)		
(2)	Percent of previous 6 days that teacher taught Utkarsh	0.826***	0.844***	298	0.28
		(0.013)	(0.013)		
(3)	Did Utkarsh worksheet on the day of the survey or on the recent most day the teacher taught	0.977***	0.943***	298	0.04
		(0.012)	(0.014)		
(4)	Implementing the correct phase	0.940***	0.965***	299	0.17
		(0.016)	(0.013)		
(5)	Students who attended FC met the inclusion criteria	0.833***	0.853***	298	0.57
		(0.028)	(0.028)		
(6)	Students who attended SLP met the inclusion criteria	0.804***	0.796***	233	0.87
		(0.039)	(0.036)		
(7)	Students who attended CC met the inclusion criteria	0.781***	0.772***	288	0.83
		(0.032)	(0.033)		
(8)	Share of all FC lessons that are Utkarsh lesson	0.851***	0.849***	298	0.93
		(0.020)	(0.023)		
(9)	Followed Utkarsh lesson exactly as instructed in the lesson guides	0.980***	0.983***	294	0.81
		(0.010)	(0.010)		
(10)	Students currently using handbooks in class (directly observed)	0.980***	0.913***	244	0.08
		(0.025)	(0.032)		
(11)	Classroom has a word wall (directly observed)	0.093**	0.123***	299	0.55
		(0.043)	(0.042)		
(12)	Student desks arranged in small group (directly observed)	0.798***	0.781***	249	0.77
		(0.051)	(0.051)		
(13)	Teaching planned lesson during FC	0.862***	0.900***	84	0.50
		(0.058)	(0.047)		
(14)	Percent of previous 6 days that teacher followed planned Utkarsh during FC	0.714***	0.796***	84	0.00
		(0.042)	(0.029)		

Notes: Table is showing school-level implementation fidelity of the Utkarsh program. All outcomes are set to zero for all control group schools. Each regression includes the following school-level variables: average teacher age and its square, average teacher experience and its square and share of female teachers. Rows 1-4: teacher reported responses measured during the spot visit. Rows 5-7: Teacher reported measure of whether students attending each phase met the inclusion criteria of that phase. These outcomes do not consider any inclusion error. Rows 5-6: measured during spot visit. Row 6 is measured only if the spot visit took place during the SLP phase. Row 7: measured during the endline. Row 8: teacher reported responses measured during the spot visit. Row 9: measured during the endline. Rows 10-12: enumerator observed outcomes measured during the spot visit. Rows 13-14: teacher reported responses measured during the spot visit and measured at all schools if spot visit took place during FC phase. Hence the sample sizes of these outcomes are less than other variables. Heteroskedasticity-robust standard errors, clustered at the school level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table A15: Headmaster Attendance

	Headmaster in school during monitoring visit
	(1)
Standard Utkarsh	0.083
	(0.055)
Flexible Utkarsh	0.066
	(0.052)
Observations	298
Control group mean	0.14
Standard Utkarsh=Flexible Utkarsh (p-value)	0.76

Notes: Regression includes strata, week, and day-of-week fixed effects; a dummy for headmaster being female, age and age squared, years of experience and years of experience squared, the school having multiple sections for class 9, total school enrollment, and an indicator variables for monitoring visit phase. Heteroskedasticity-robust standard errors, clustered at the school level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table A16: Subject Teacher Surveyed during Monitoring Visit Survey

We surveyed at least one .... teacher	Every Subject	English	Math	Odia	Science
	(1)	(2)	(3)	(4)	(5)
Flexible Utkarsh	0.089	0.088	0.087	0.083	-0.104
	(0.066)	(0.056)	(0.058)	(0.059)	(0.064)
Observations	200	200	200	200	200
Standard Utkarsh Mean	0.36	0.77	0.78	0.74	0.81

Notes: Each column is a school-level observation showing whether teacher of an Utkarsh subject was interviewed during the monitoring visit survey. All regressions include following school-level variables: average teacher age and its square, average teacher experience and its sqaure and share of female teachers. Column 1: shows whether at least one teacher of every subject was surveyed. Columns 2-5: show whether subject teacher of the respective subject was surveyed. Heteroskedasticity-robust standard errors, clustered at the school level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table A17: Teacher Flexibility and Discretion

	(1) Filled out flexible Utkarsh teaching plan	(2) Followed flexible Utkarsh teaching plan for this week	(3) Had autonomy in using Utkarsh	(4) Teacher can adjust content if students have difficulty
Standard Utkarsh	0.017	0.014	0.945***	0.009
	(0.018)	(0.016)	(0.013)	(0.030)
Flexible Utkarsh	0.200***	0.139***	0.989***	0.060*
	(0.034)	(0.028)	(0.010)	(0.031)
Female (=1)	-0.010	0.002	-0.008	-0.015
	(0.029)	(0.025)	(0.016)	(0.041)
Age of the teacher (in years)	-0.031**	-0.024*	0.002	0.022
	(0.013)	(0.013)	(0.007)	(0.020)
Age squared	0.000**	0.000*	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Teaching experience (years)	0.013**	0.009**	-0.005	-0.012
	(0.006)	(0.005)	(0.003)	(0.011)
Teaching experience squared	-0.000**	-0.000**	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Subject teaches (=1)				
English	0.011	-0.005	0.031	0.032
	(0.048)	(0.039)	(0.025)	(0.065)
Odia	-0.007	-0.021	-0.001	-0.083
	(0.051)	(0.043)	(0.028)	(0.064)
Math1	-0.013	-0.039	-0.027	0.015
	(0.041)	(0.039)	(0.026)	(0.054)
Math2	0.041	0.019	-0.021	-0.044
	(0.039)	(0.035)	(0.029)	(0.056)
Science 1	0.010	-0.024	0.023	0.072
	(0.029)	(0.031)	(0.025)	(0.054)
Science 2	-0.029	-0.038	0.022	-0.048
	(0.037)	(0.030)	(0.026)	(0.047)
History	-0.066*	-0.077***	0.019	-0.029
	(0.035)	(0.029)	(0.025)	(0.060)
Geography	-0.025	-0.030	-0.015	-0.003
	(0.036)	(0.043)	(0.026)	(0.055)
Hindi	-0.072	-0.104**	0.022	0.289***
	(0.053)	(0.050)	(0.039)	(0.103)
Sanskrit	0.064	0.039	0.025	0.025
	(0.088)	(0.091)	(0.025)	(0.115)
Work Experience	-0.123	-0.163	0.022	-0.758***
	(0.128)	(0.118)	(0.039)	(0.093)
EVS	-0.216**	-0.183**	-0.009	0.181
	(0.093)	(0.079)	(0.038)	(0.116)
Other			0.072	0.093
			(0.055)	(0.166)
Has a teaching certificate (=1)	0.014	-0.001	-0.036*	0.068
	(0.029)	(0.029)	(0.019)	(0.042)
Days absent from work	-0.037***	-0.031**	0.000	0.006
	(0.014)	(0.013)	(0.008)	(0.021)
Works in another school(=1)	-0.045	-0.035	0.050	-0.231*
	(0.091)	(0.049)	(0.047)	(0.131)
Time spent in preparing for lesson (hours/week)	-0.002	-0.002	-0.001	0.000
	(0.002)	(0.001)	(0.001)	(0.002)
Time spent grading (hours/week)	0.001	-0.001	-0.001	-0.001
	(0.003)	(0.002)	(0.001)	(0.002)
Select teaching materials, methods, strategies (=1)	0.069*	0.032	0.039	-0.052
	(0.035)	(0.041)	(0.052)	(0.084)
Allowed to modify course timetable (=1)	-0.038	-0.055	0.015	0.233***
	(0.035)	(0.034)	(0.022)	(0.059)
Others do not select evaluation activities (=1)	-0.074**	-0.049	-0.029*	0.019
	(0.031)	(0.031)	(0.016)	(0.041)
Feel pressure to complete curriculum during school year (=1)	-0.053*	-0.029	0.013	0.024
	(0.028)	(0.027)	(0.019)	(0.041)

...(continue in next page)...

Table A17: Teacher Flexibility and Discretion (continue)

	(1) Filled out flexible Utkarsh teaching plan	(2) Followed flexible Utkarsh teaching plan for this week	(3) Had autonomy in using Utkarsh	(4) Teacher can adjust content if students have difficulty
Mean autonomy (standardized)	-0.000	-0.012	0.006	-0.016
	(0.015)	(0.018)	(0.010)	(0.023)
According to the teacher, percent of students who				
Will pass board exam in first try	0.000	0.001	0.000	-0.001
	(0.001)	(0.001)	(0.001)	(0.002)
Can write a simple English sentence	0.001	0.001	-0.000	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Can do a three-digit sum	0.000	0.000	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)
Will eventually pass the class 10 board exams	-0.001	-0.001	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.002)
Will eventually pass the class 12 board exams	-0.001	-0.001	0.001	0.000
	(0.001)	(0.001)	(0.001)	(0.002)
Will eventually complete bachelors	0.000	-0.000	-0.001	-0.002
	(0.001)	(0.001)	(0.001)	(0.002)
Teacher feels (=1)				
Curriculum should be followed even if students have lower learning levels	0.049**	0.053*	-0.002	0.083**
	(0.023)	(0.029)	(0.015)	(0.041)
That if students are not ready for board exams, it would be the teacher's own fault	-0.001	-0.025	-0.013	0.006
	(0.025)	(0.023)	(0.019)	(0.040)
Valued and appreciated	-0.008	0.002	0.041*	0.021
	(0.030)	(0.026)	(0.021)	(0.042)
Satisfied with job	-0.019	-0.027	0.045	0.033
	(0.033)	(0.026)	(0.028)	(0.051)
That their opinion seems to count	-0.041	-0.053	0.018	0.155**
	(0.042)	(0.042)	(0.031)	(0.071)
That they have the materials and equipment to teach effectively	0.026	0.046*	-0.011	-0.003
	(0.024)	(0.025)	(0.019)	(0.042)
Similarly or more effective compared to colleagues	0.059	0.073	0.056	-0.067
	(0.052)	(0.071)	(0.047)	(0.091)
Observations	569	569	834	834
Control group mean	0.00	0.00	0.00	0.77
Standard Utkarsh = Flexible Utkarsh (p-value)	0.00	0.00	0.00	0.09
Source	Directly Observed	Directly Observed	Self-reported	Self-reported

Notes: All regressions include strata, week, and day-of-week fixed effects. Columns 1-2, measured at midline, include indicator variables for monitoring visit phase. Column 1 measures whether the teaching plan is filled out. Column 2 is a self-reported measure of whether teachers followed the teaching plan that they filled out for that week. Columns 3-4, measured at endline, includes an indicator variable for early endline visit. Columns 3 is a self-reported measure of teaching having autonomy in using Utkarsh lessons. Column 4 is a self-reported measure of whether teacher can adjust content of the lesson. Heteroskedasticity-robust standard errors, clustered at the school level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table A18: Teacher Flexibility and Discretion Heterogeneity by BL Student Test Score Standard Deviation

	Filled out flexible Utkarsh teaching plan	Followed flexible Utkarsh teaching plan for this week	Had autonomy in using Utkarsh	Teacher can adjust content if students have difficulty
	(1)	(2)	(3)	(4)
Standard Utkarsh	0.037	0.109	0.936***	-0.191
	(0.111)	(0.098)	(0.068)	(0.162)
Standard Utkarsh*BL Test Score SD	-0.020	-0.103	0.009	0.214
	(0.120)	(0.106)	(0.073)	(0.175)
Flexible Utkarsh	-0.012	0.068	0.924***	-0.332**
	(0.165)	(0.127)	(0.053)	(0.164)
Flexible Utkarsh*BL Test Score SD	0.239	0.092	0.070	0.441**
	(0.178)	(0.134)	(0.058)	(0.177)
BL Test Score SD	-0.025	0.041	-0.045	-0.218*
	(0.076)	(0.062)	(0.032)	(0.125)
Observations	569	569	834	834
Control group mean	0.00	0.00	0.00	0.77
Standard Utkarsh+Standard Utkarsh*BL Test Score SD=0 (p-value) (p-value)	0.42	0.71	0.00	0.49
Flexible Utkarsh+Flexible Utkarsh*BL Test Score SD=0 (p-value) (p-value)	0.00	0.00	0.00	0.00
Source	Directly Observed	Self-reported	Self-reported	Self-reported

Notes: All regressions include strata, week, and day-of-week fixed effects, teacher's age in years and age squared, teacher's years of experience and experience squared, a dummy for teacher being female, and a vector of dummy variables for main subject taught by teacher in the baseline. Columns 1-2, measured at midline, include indicator variables for monitoring visit phase. Column 1 measures whether the teaching plan is filled out. Column 2 is a self-reported measure of whether teachers followed the teaching plan that they filled out for that week. Columns 3-4, measured at endline, includes an indicator variable for early endline visit. Columns 3 is a self-reported measure of teaching having autonomy in using Utkarsh lessons. Column 4 is a self-reported measure of whether teacher can adjust content of the lesson. Heteroskedasticity-robust standard errors, clustered at the school level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table A19: Additional Teacher Outcomes

	Burnout index	Stress index	Anxiety index	Lesson preparation time (hours/week)	Grading time (hours/week)	Teacher enjoyed Utkarsh
	(1)	(2)	(3)	(4)	(5)	(6)
Standard Utkarsh	0.105	0.059	0.025	-0.229	0.410	0.953***
	(0.089)	(0.082)	(0.090)	(0.625)	(0.558)	(0.014)
Flexible Utkarsh	0.135	0.082	0.049	-0.842	0.259	0.940***
	(0.089)	(0.090)	(0.096)	(0.649)	(0.551)	(0.016)
Observations	834	834	834	834	834	834
Control group mean	0.00	0.00	0.00	13.65	8.94	0.00
Standard Utkarsh = Flexible Utkarsh (p-value)	0.75	0.79	0.79	0.34	0.78	0.54

Notes: All regressions include strata, week, and day-of-week fixed effects; teacher's age in years and age squared, years of experience and experience squared, a dummy for being female, a vector of dummy variables for main subject taught in the baseline, and indicator variables for early endline visit. Column 1 is an inverted covariance matrix weighted standardized index generated following Anderson (2008) from the following variables: teacher feeling mentally exhausted from work; feeling fatugied; feeling having a positive influence on people; feeling very energetic about job; feeling satisfied with job at school. Columns 2 and 3: measured on the Depression Anxiety Stress Scale (DASS). To construct respectived indices, response to relevant questions of the scale are summed and then standardized. Heteroskedasticity-robust standard errors, clustered at the school level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table A20: Teacher Competency

	Self-Assessed Effectiveness			Competency Exam				
	At least as Effective as Other Teachers	More Effective Than Other Teachers	Much More Effective Than Other Teachers	Test Score (Percent)		Competency Exam Administered in School (=1)	Subject Teacher Tested in the Competency Exam Was Surveyed at Baseline (=1)	
				English	Math		English	Math
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Standard Utkarsh	0.012	-0.005	-0.007	-0.476	1.067	0.009	0.069	-0.008
	(0.012)	(0.041)	(0.021)	(2.399)	(1.591)	(0.019)	(0.072)	(0.056)
Flexible Utkarsh	0.003	0.020	-0.021	-0.943	2.909**	-0.015	0.050	-0.078
	(0.013)	(0.040)	(0.020)	(2.655)	(1.413)	(0.023)	(0.076)	(0.058)
Observations	834	834	834	226	303	300	226	303
Control mean	0.97	0.41	0.08	57.83	89.21	0.80	0.65	0.64
Standard Utkarsh=Flexible Utkarsh (p-value)	0.39	0.52	0.51	0.83	0.24	0.24	0.76	0.24

Notes: All regressions include strata fixed effects. Columns 1-5: Regressions include week-of-survey, and day-of-week fixed effects, a dummy for teacher being female, age of the teacher and age squared, teacher's years and experience and experience squared. Columns 4-5 measure test scores test scores in respective subjects by respective subject teachers on the teacher competency examination. Regressions include a vector of dummy variables for main subject taught in the baseline, and indicator variables for early endline visit. Column 6: measures that the school participated in the teacher competency examination. Regression includes a dummy for headmaster being female, headmaster's age and age squared, headmaster's years of experience and years of experience squared, school having multiple class 9 sections, and total enrollment. Columns 7-8: indicator variables measuring whether the respective subject teacher was surveyed at baseline. Regressions include week-of-survey, and day-of-week fixed effects, a dummy for teacher being female, age of the teacher and age squared, teacher's years and experience and experience squared. Heteroskedasticity-robust standard errors, clustered at the school level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table A21: Longer-Term Follow-up Survey Details

	Board exam scores of our study sample collected (1)	Analytical Sample Students Surveyed in the Longer-Term Follow-up Survey (2)	Student took offline exam (3)
Standard Utkarsh	-0.010	0.019	0.018
	(0.020)	(0.013)	(0.014)
Flexible Utkarsh	0.021	0.008	0.011
	(0.019)	(0.014)	(0.015)
Observations	5,756	5,457	1,253
Control group mean	0.92	0.22	0.05
Standard Utkarsh=Flexible Utkarsh (p-value)	0.13	0.41	0.64

Notes: This table contains data from a longer-term follow up survey conducted in November-December 2021. All regressions include strata fixed effects; student's standardized IRT scores from baseline English, Math, and Odia tests, a dummy for student being female, and age of student. Heteroskedasticity-robust standard errors, clustered at the school level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table A22: Detailed Effects on Board Exam Marks

	Grades			Test Score in Percentage			
	A	C or above	D or above	English	Math	Odia	Science
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Standard Utkarsh	-0.011**	-0.069***	-0.046**	-0.123**	-0.122**	-0.123***	-0.136***
	(0.005)	(0.022)	(0.018)	(0.051)	(0.056)	(0.044)	(0.051)
Flexible Utkarsh	-0.002	-0.046**	-0.028	-0.094*	-0.045	-0.079*	-0.087
	(0.006)	(0.022)	(0.018)	(0.053)	(0.053)	(0.046)	(0.054)
Observations	18,551	18,551	18,551	18,551	18,551	18,551	18,551
Control group mean	0.05	0.65	0.85	0.00	0.00	0.00	0.00
Standard Utkarsh=Flexible Utkarsh (p-value)	0.15	0.33	0.35	0.60	0.17	0.36	0.37
Raw control mean				0.53	0.58	0.56	0.55
Raw control SD				0.15	0.16	0.16	0.15

Notes: All regressions include strata fixed effects; standardized IRT scores from baseline English, Math, and Odia tests, a dummy for student being female, and age of student. All outcomes are from administrative data on board exam results. Columns 1-3: indicating whether the student achieved the respective grade or above. Columns 4-7 are total scores in respective subjects. Colum 5 is. Heteroskedasticity-robust standard errors, clustered at the school level, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

A Additional Implementation Details

The program begins by holding a one-week training session for all schools in the program immediately before the beginning of the school year. School headmasters and a one teacher for each of the four targeted subjects are invited, and the training centers around how to use handbooks that explain how to implement the program Utkarsh subject-specific handbooks into an effective teaching practice. All program schools receive teaching and learning materials developed by PFA, which include the teacher handbooks as well as student handbooks and workbooks; the workbooks have worksheets for the students to complete for each day of the program. PFA helps to run the training sessions and collaborates with the government to monitor implementation and maintain quality. In Odisha, the partner government department is the Department of School and Mass Education (SME). For the version of the program we study in this paper, PFA conducted the training sessions themselves. The program was also scaled up in the rest of Odisha starting in the 2019-20 school year; for this broader scale-up, PFA used a cascade (train-the-trainers) model to run the trainings, teaching SME staff how to do the actual training of teachers .

B Additional Test Construction Details

We used bespoke exams to test student learning. We constructed the tests specifically for evaluation and did not share them with PFA, SME, or any other entity involved in the implementation of Utkarsh during the period of evaluation. Our test questions were based upon learning objectives from the official curriculum and covered material from Class 3 through Class 9. For English and math tests, we included questions from PISA (four English questions and six math questions) that may not necessarily map to the official curriculum.

We used item response theory (IRT) to construct students' test scores, pooling all questions from baseline and follow-up. The IRT scores for math, English, and Odia are standardized by the baseline mean and standard deviation. Since we administered the science test

only at the follow-up, we use the control group mean and standard deviation to standardized the science test score. We also construct an overall score using IRT by pooling all baseline and follow-up questions of all subjects. We then standardize with respect to baseline mean and standard deviation.

We calculated the students' grade-level mastery, competency level, and number of grades behind in English, math, and Odia based on their ability to respond to specific grade-level questions correctly.[19]

[19]Students have Class 8 mastery if they correctly answered at least 75 percent of Class 8 level questions. Students had Class 5 level mastery if they answered at least 75 percent of Class 5 level questions correctly but less than 75 percent of Class 8 level questions. Similarly, a student had grade 3 level mastery if they correctly answered at least 75 percent of Class 3 level questions but less than 75 percent of Class 5 questions. Students who answered fewer than 75 percent of class 3 questions correctly are considered to have class 2 mastery. We construct grades behind by subtracting grade level competency from 8. For instance, a student with competency at grade level 8 is 0 grades behind, while a student with below grade 3 level competency is 6 grades behind.